# DOCUMENT RESUME

ED 071 729 LI 004 133

| | |
|---|---|
| AUTHOR | Penland, Patrick R. |
| TITLE | Communication Research for Librarians. |
| INSTITUTION | Pittsburgh Univ., Pa. Graduate School of Library and Information Sciences. |
| PUB DATE | 72 |
| NOTE | 132p.; (72 References); Discourse Units in Human Communication For Librarians |
| AVAILABLE FROM | Bookstore, University of Pittsburgh, Pittsburgh, Pa. 15213 ($3.50) |
| EDRS PRICE | MF-$0.65 HC-$6.58 |
| DESCRIPTORS | *Communication (Thought Transfer); Definitions; Information Science; *Librarians; Library Science; Research Design; *Research Methodology |
| IDENTIFIERS | *Librarianship |

ABSTRACT

Research design is a product of the scientific method in Western Civilization whose major purposes are to eliminate the biased judgements of individual researchers and to ensure replication of the study. In general, the formal protocol of the research design includes several elements: theoretical position out of which the hypotheses to be tested grew; relation of definitions to the measurement scales; methods of observation (data collection) and sampling; analysis of data and summary of findings; conclusions and recommendations for further research. Such in outline is the overall format of the formal report of basic research. With it alone, another investigator should be able to perform precisely the same study with no other information but the protocal of the research design itself. The task of an author of a research study is to establish, and then describe with precision, the way in which the following elements of a study are to be defined: subject, attribute, setting, moment, and method. (For related documents see: LI 004 134, ED 048 902, ED 049 801 through 049 804, and ED 054 840-054 841) (Author/NH)

ED 071729

U.S. DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

# COMMUNICATION RESEARCH
# FOR LIBRARIANS

Patrick R. Penland
Graduate School of Library and Information Sciences

LI 004 133

1972
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

(Discourse Units in Human Communications for Librarians)

# Table of Contents

Proposal Development..............................1

Primer on Theory..............................15

Measurement Design..............................29

Observational Methods..............................46

Experimental Control and Extrapolation..........71

Search and Discovery..............................89

Glossary..............................109

General References..............................125

# PROPOSAL DEVELOPMENT

Every individual has a need for reflection about his tasks and his environment. But it is the professional person who develops an attitude of planned reflection in order to go beyond unthinking activities determined by tradition and emotion. The professional person in information, library and media science is the individual in the social group willing to recognize problem situations and work for a solution. A research study is an opportunity for the professional person to demonstrate the ability to do reflective thinking in his field of specialization, and to organize the productive results thereof into a research paper. A research study may be undertaken to fill a gap in knowledge (e.g. experimental study), or to supply more intensive documentation (e.g. case studies) in a discovery context.

The eventual development of a research study is predicated upon the existence of a felt need or dissatisfaction with the status quo in the professional person's field of specialization. Something about his field of study either may not make sense to him or else is the occasion of a major discrepancy between one principle and another or between theory and practice. A felt difficulty in his area of professional concern may occur for a variety of reasons. The means employed in a particular situation may not be appropriately adopted to an end or objective in communication services. In other words, a course of action has been developed without reference to, or an inappropriate reference to purpose. Consequently the character and structure of the development have not been clearly identified. In another instance, an unexpected event may occur which has not been or was not able to be explained satisfactorily in terms of communication principles and theory.

As a result of this feeling of intellectual uneasiness, the professional observer has read a great deal of the literature in his area of concern. He has examined the available research studies, the articles on trends and developments, as well as the statements of field experience. Nevertheless his feeling of uneasiness persists and indeed is strengthened because of his reading. Apparently, in the literature, other professional persons have either noted this concern or are at least aware of it. However, the literature survey has indicated that no one has developed the concern into a formal problem statement together with an hypothesis for solution; or if they have, the problem has not been given the attention it should have. Consequently the professional observer (researcher) can justifiably regard this concern as a problematic situation for his own research proposal.

Now that a problematic situation has been identified the observer (researcher) finds that he can begin to reflect upon the concern and determine why it has occured. At the same time insight begins to develop as a possible solutions. In this process the observer (researcher) will again survey the professional literature and any previous research done in this particular area of concern. The insight which has developed becomes an organizing influence for his reading and thinking. At this point it is well to develop and utilize a rough outline. Such an outline should not be arbitrarily imposed. It should grow out of the researcher's study and reflection in an organic way, giving evidence that some thinking has taken place.

In the process of developing a research proposal, it is evident that the concepts of communications science, thinking and research are almost synonymous. The activities associated with each of these concepts are natural to the ordered processes of mind of the normal person. A feeling of doubt has occured in an area of professional concern and has provoked a consequent need for certainty. Under the drive for psychological closure, a definite question has been isolated and a tentative solution (insight) selected for investigation. Under investigation (proposal development), this solution, or some better solution, is eventually accepted in light of the evidence obtained. At this point psychological closure (homeostasis) is complete and the need for certainty satisfied. However as the habit of scientific inquiry and reflective thinking becomes more ingrained, the communications researcher begins to realize that the search for certainty is an illusion or, at best, a provisional pause upon a plateau before further necessary research becomes evident.

Once the preliminary and basic reflective thought processes have been stimulated into productivity, the communicator is in a position to give attention to the formal outline of his proposal. Some manual of style should be accepted and followed consistently. An academic institution, advising students through a thesis for example, may indicate what manual is acceptable. For example, doctoral students at the Graduate School of Library and Information Science, University of Pittsburgh use the Manual of Style prepared by Martha Manheimer (New York, Dekker, 1972).

The first requirement of a formal outline is a short and pertinent title. A title is in brief an abstract of, or an index to, a thesis, a monograph, or any work of the human mind. If greater specificity is required, a sub-title may be employed. The communicator should keep in mind that the title may be the only entry the study will have in an index to research or in the library card catalog. In fact any writer (recorder) should appreciate how few readers will thumb through all of the titles beginning "History of..."; "Relationship between..."; etc. It is to the writer's advantage to compress the research title to its barest essential words.

Next in order after the title is the statement of purpose. The purpose should be stated in paragraph form directly below the title. The statement of purpose in the formal outline is simply a formulation in carefully chosen words of the original intuitive insight that occured to the communicator when reflecting upon the problematic concern in his area of specialization. As his reading and thinking progressed, this insight grew into a conviction and became the organizing influence in assembling data to substantiate the hypothetical guess. The intuitive guess becomes the hypothesis, or purpose, of the thesis. In order to substantiate (or perhaps indeed negate) the hypothesis, further research is conducted.

Whether the hypothesis is substantiated, modified, or simply negated through appropriate research methods, the proposal as a "thesis" may be a valid piece of work for a pilot investigation. Consequently it is obvious that methodology plays a considerable role in thesis development. Before considering methodology however, some attention must be given to a justification of the research topic. In general, justification for the selection of the problem grows out of the reasons listed for its personal suitability and its general value. The research study is a professional contribution. This means, in short, that the original problematic situation with which the observer (researcher) felt uncomfortable or uncertain, becomes the basis for the justification.

These preliminary comments should not be taken as sanctioning a subjective approach to research development. Nevertheless, the whole process of research design is to objectify hypothetical insights, to relate them to, and to validate them by objective professional knowledge. It is the function of a communicator to identify professional and especially social problems. That he has identified a problem is surely an accomplishment. Problem identification indicates that the researcher (observer has been reflecting productively upon the principles and practices of communication and has been attempting to integrate his observations into some professional area of concern. Actually, to develop a communications research proposal is to demonstrate that the specialist in information, library and media science is able to meet the minimum requirements for a professional person in society.

## Research Development:

Research design is a product of the scientific method in Western Civilization whose major purposes are to eliminate the biased judgements of individual researchers and to ensure replication of the study. In general, the formal protocol of the research design includes several elements: theoretical position out of which the hypotheses to be tested grew; relation of definitions to the measurement scales; methods of observation (data collection) and sampling; analysis of data and summary of findings; conclusions and recommendations for further research. Such in outline is the overall format of the formal report of basic research. With it alone, another investigator should be able to perform precisely the same study with no other information but the protocal of the research design itself.

Most textbooks in research methodology proceed in this formal manner since it is difficult if not impossible to represent the dynamic and often cybernetic pattern of research thinking. Indeed, if research thinking is anything, it is a dynamic intrapersonal development including both deductive and inductive processes. One or the other of these thought processes may predominate at any point in the researcher's intellectual development.

Neglected in many approaches to the study of research design is the inductive, and indeed largely pragmatic manner in which research proposals are developed. This is particularly the case in areas where neither the theoretical constructs nor measurement scales have been developed to any great extent. As the researcher continues to observe subjects and develop measurement scales, his thought begins to expand in two directions: theoretical implications of his measurement scales; statistical manipulations of the data to be obtained for justifying theoretical implications of the hypotheses. The task of an author of a research study is to establish, and then describe with precision, the way in which the following elements of a study are to be defined: subject, attribute, setting, moment, method.

Subject is the entity about which an observation may be made and to which a theoretical construct can be imputed. Subject does not necessarily mean a subdivision of knowledge. It is a specific entity upon which an observation is to be made. This particular entity must be clearly distinguished from all other apparently similar entities. "A freshman student at the University of Pittsburgh School of Library Science" is an example of a subject which is fairly specifically identified. "A student" is not.

The identification of a specific subject involves a description of some of its attributes. The creation of groups or classes is the result of observing similar attributes in a series of independent entities. A subject is seldom observed as a totality, *e.g.*, a subject heading or library service. Rather a specific characteristic is observed, an attribute, trait, quality, or property. A patron as such is not observed, but only his height, his color, his behavior, his structure, that is the specific property or attribute to be subjected to observation. If several attributes are to be observed, each in itself involves a separate observation and each in reality constitutes a separate study.

Attribute is the property or characteristic of the subject that is to be observed and explained eventually by a theoretical construct. An attribute is observed in a number of ways. The observer may use any of his senses, or any written, mechanical, electrical, or chemical means by which to identify or measure an attribute.

The observation of an attribute or property is usually made in relation to some standard, frame of reference, or unit of measure validated upon a theoretical construct. In describing an attribute being observed it is necessary to specify it with reference to the standard being employed. For example, the following observations are specifically defined: "non-fiction books in Dewey 700 class;" "semantic differential test." And the following are not: "library use;" "intelligence."

Setting includes the conditions under which the attribute is to be observed. Theoretically, any aspect of the physical, or the biological or social environment may affect an attribute or an observation made of it. It is essential to identify the circumstances which obtain at the time the subject of a study is undergoing the specific observation.

In the physical sciences, a simulated or artificial environment may be created and mechanical or electrical devices employed in order to minimize human error in making observation. But even when "control" is possible the environment so created must be specified. In many instances, however, the observer may have to accept the subject in whatever circumstances it can be found and make his observations within that environment. An objective view of the environment must be maintained in order to identify those features which will affect the observation. An example of specified conditions could be: "non-fiction books in Dewey 700 class borrowed by patrons in the age group 18-20 during August." An example of an insufficiently specified condition is: "measurements of library adult education."

Moment is the time period in which an observation is to be made. Many observations can only be made at some point in time after an occurence, and not on a supposition, e.g., "I wonder what would happen if..." This dichotomy is the classical distinction between the survey and the scientific method. Of course techniques can be refined and records or measures developed in order to reduce time lag. It may appear as if the observation is made simultaneously with the happening of an event, but in reality the record is made only after the fact. What happens at a specific moment is assumed to have taken place on the basis of subsequent observation.

It makes considerable difference at what point in time the individual makes an observation. Time must always be specified. The degree to which time is specified is dependent on the type of observation made. In each study there is a unit (whether a micro-second or a year) which makes the specific observations meaningful. Time is specified in the following example: "library-user questionnaire administered during the first week of August." Time is not specified in this example: "intelligence on beginning to use the library."

Method is the means by which data on the attribute is to be obtained and manipulated statistically. With the repetition of observation, the observer may begin to relate a series of findings one to another and thus commence the process of inductive reasoning from those findings. The method of a research study is to note the consistency or inconsistency of repeated observations. The attempt to reason from these observations is almost an independent activity. At least it is more akin to the formal presentation of a research design than it is to considerations of measurement and data collection.

If an individual makes two observations (or two sets of observations) of an attribute and finds a difference, one or more elements must have become modified: (1) subject, (2) the observational method, (3) the time when observation is made, (4) or the conditions under which observation is made. The essential problem is to identify which modifications have occured and the contribution played by each. In general this may be attempted by two methods:

> Begin with the fact that a difference has been observed and examine each factor that might have contributed to the difference. This is the analytical approach. For example, the birth rate in the United States may be observed to be different in 1936 as compared to a similar observation in 1955. An observer than attempts to determine the various factors which were in operation at the two points in time in order to account for the observed difference.
>
> Begin with a set of factors and by altering one examine the accompanying observational difference. This is an experiment. As an example, an investigator might present a lecture on research methods to a group of students. He would subsequently observe the ability of the individuals to prepare a research topic as opposed to a similar group of individuals who were not exposed to the lecture.

Data and Design:

In most research courses, measurement if dealt with at all is done in a cursory manner. From a formal point of view it may be of secondary importance. However from a functional and "how-to-do-it" viewpoint, the role and characteristics of measurement cannot be avoided as well as its direct relationship to statistical analysis. There are four major types of measurement scales: nominal, ordinal, interval and ratio. In most instances, the scale employed will have to have interval characteristics in order that statistical manipulations may be employed on the data. In general, there are four patterns in the organization of observations and data collection (see figure -- Organization of Observations):

> Repeated observations of a specific type on a specific subject to identify a specific and consistent characteristic.
>
> Repeated observations of several types to identify distinguishing characteristics.
>
> Repeated observations of several types to demonstrate an association of specific characteristics, i.e., examination of associations between characteristics.
>
> Repeated observations of a characteristic before and after the introduction of a specific modification, i.e., demonstrating the modification of a characteristic.

Whenever the observer has two subjects, they are, by definition, different. They are made up of separate units of matter. They may, of course, be quite similar in certain characteristics. Differences in properties or traits may not at first be detectable by a particular method of observation. The observer may even consider them identical. But any such statement should be recognized as an assumption.

A particular observational method may not reveal difference between two or more subjects. As such, the observation is assumed to reveal one class or group. A group is but a series of individual subjects wich can be identified as having a particular characteristic by means of a particular observational method.

The subdivision of a group into subclassifications (as in Dewey classes) is accomplished by identifying a particular observation consistency or characteristics in a series of individual. Any other observation consistency can be logically only some portion of these individual. Subdivision of a group may also be desired in order that

## ORGANIZATION OF OBSERVATIONS

| Study Design | Title | Objective or Hypothesis |
|---|---|---|
| Repeated observations of a specific type on a specific subject to identify a specific and consistent characteristic. | Measurement of Circulation Loan Size of Young Adults, Ages 18-20. | To determine the number of books (i.e. more then 80 pages) in the hundred Dewey classes borrowed by male young adults between 18-20 years of age. |
| Repeated observations of several types to identify distinguishing characteristics. | Examination of Size and Number of Loans to Distinguish Reading Patterns. | To determine the size, composition and frequency of loans of library materials in relation to stated reading interests of young adults between 18-20 years of age. |
| Repeated observations of several types to demonstrate an association of specific characteristics, i;e. examination of associations between characteristics. | Association of Smoking and Academic Performance in Library School Students. | Students with high academic performance in a particular school year will be shown to have differences in smoking habits during the year when compared with students with low performance. |
| Repeated observations of a characteristic before and after the introduction of a specific modification, i.e. demonstrating modification of a characteristic. | The Effect of Extended Loan Period on Reading Patterns. | The application of a longer loan period (4 weeks instead of 2) with automatic renewal will result in reading patterns that show less subject dispersion. |

two or more similar groups can be observed separately. The essence of such subdivision is an allocation of individuals so that a consistency in characteristics exists from individual to individual not only within each group, but also between groups.

The observer may modify any of the elements of the study and make two or more observations on a series of individual subjects. He might wish to make identical observations of the of the same attribute on various subjects. If, for example, he were interested in obtaining information on the smoking habits of library students, he might select two subgroups: "Pitt library students and Chicago library students;" or, "Pitt master's degree library students, and Pitt post-master's library students."

The observer merely repeats the same observation on each of the individuals. He identifies those who display specific attributes of form and quantity of tobacco consumption. The observation may be a difference in smoking habits of one group vs. the other. Actually what has been done is to identify a number of individuals having a specific habit of tobacco usage within each of the groups studied at a specific moment in time.

The observer can modify time, manner of observation and conditions. When he modifies time, he observes the same group of library students at two points of time, eg. on the first day of classes in the beginning year and on the first day of classes in the advanced year. He might note a difference in the individuals at these two points in time.

He might also modify the form of observation and note the information elicited by interview as opposed to a questionnaire. He could change conditions and attempt to elicit information (immediately following a lecture) on the association of cigarette smoking and lung cancer as opposed to making the observation during a fraternity smoker.

The observer has identified the frequency with which individual members of a group display a characteristic under certain factors which exist. If a difference is found the observer may wonder why, but as such this is a question of relevance to a theory. One first determines what observations are obtained and whether there is a difference, before one can concern himself with the matter of accounting for a difference by hypothesizing about it.

The discovery and testing of causal relations is an important and actually a very basic aspect of any data collection methods. The process involves the approach of inductive logic usually based on Mill's canons of inquiry. However, the process presupposes some reason analysis or validation i.e. previous knowledge and conditions which control the application of the canons. The first steps in reason analysis include the following:

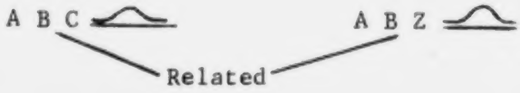
> Questions must be formulated with sufficient definiteness so as to guide the research style i.e. hypotheses development.
>
> Complex conditions which the investigator finds so overwhelming at the start must be "boiled down" into component parts, i.e. assumption identification.
>
> All relevant conditions must be constantly considered. Relevant factors can be varied only one at a time. Scrutiny must be exercised so as not to ignore factors which may be relevant to the results.
>
> Factors considered to be irrelevant must also be varied as much as possible in order to determine the reality of their supposed irrelevancy.

Mill's canons (rules) may be considered as methods for discovering and testing causal relations. Three of the canons are illustrated in the figure -- Variable Relationships. Originally stated as five propositions for the guidance of inductive logical thinking, the canons have since been variously combined and related in two, three or in four propositions. The original five canons were:

1. Method of Concomitant Variation.
2. Method of Differences
3. Method of Agreement
4. Joint Method of Agreement and Difference
5. Method of Residues

## VARIABLE RELATIONSHIPS

### Concomitant Variation:



If a change in the amount of one variable is accompanied by a comparable change in the amount of another variable in two or more cases, and the latter change does not occur in the absence of the first change, one change is the cause or the effect of the other.

Descriptive Methods: Variable A ← Is meaningfully related to → Variable B

e.g. Educational level tends to affect the amount of reading of an individual (Discovery)

---

### Method of Difference:

Situation #1 A B C → Z ( E group )
Situation #2 A B not C → not Z ( C group )

Therefore C → Z

Experimental Method: Variable A — Causes → Variable B

e.g. The use of plastic book jackets in a browsing room results in increased circulation (Discovery)

---

### Method of Agreement:

Situation #1 A B C → Z
#2 C D E → Z
#3 C F G → Z
#n C etc. → Z

Therefore C → Z

The Method of Difference and Method of Concomitant Variation are generally considered the most satisfactory of the methods. The Method of Agreement is considered to be the least satisfactory. The Method of Residues is considered to be actually but a modification of the Method of Difference. The Joint Method of Agreement and Difference is really the Method of Agreement supplemented by the study of some negative instances in order to approximate the Method of Difference. One of the most concise and comprehensive formulations and explanations of Mill's Canons can be found in. Abraham Wolf, "Scientific Method", Encyclopedia Britannica. N.Y., Encyclopedia Britannica, 1929.

Research Proposal:

While the mental gestation of a research design may be dynamic and inductive, the elements of a formally developed research proposal follow a more or less standard format. The final report will of course follow this format which is a rewriting almost in reverse of the initial gestation period. In this more formal approach to the initial phases of a research proposal, the observer's (researcher's) progress is facilitated if he proceeds in an orderly fashion in following up on the preliminary motivation. The principal components of an orderly progression for an initial research protocal can be outlined under the following headings: title, general problem, specific objective or hypothesis, identification and selection of subjects, procedures for making and recording observations, organization of the findings, projected analyses.

The title of the study is not an empty form. It should be a brief, concise statement of the major objective. It should identify the type of study to be performed. In general terms it indicates the subjects and the type of observations to be made. An example of a title for each of the four types of observational methods indicated in the figure Organization of Observations include the following:

> "Measurement of Circulation Loan Size of Young Adults, Ages 18-20"
>
> "Examination of Size and Number of Loans to Distinguish Reading Patterns".
>
> "Association of Smoking and Academic Performance in Library School Students".
>
> "The Effect of Extended Loan Period on Reading Patterns".

The statement of the general problem is in the nature of an introduction to the research proposal. It delineates the area of interest of the study, and includes a review of the existing literature, i.e.

the state of the art review. It places the proposed study in the context of a problem and provides a justification for entertaining the hypothesis. In effect, this statement is a logical validation of the argumentative reasoning which constitutes the research design.

The statement of the specific objectives or hypotheses will of course eventually require a formal exposition of the cause-effect relationships to be submitted to proof. However, in the context of proposal development, the hypotheses constitute simply the objective of the study. A study is supposed to result in gathering information that bears on the truth of a specific logical proposition. Consequently, the objective of a study is often expressed in the form of an affirmative proposition. When dealing with purely descriptive rather than analytical types of studies, however, it may be artificial and awkward to express the objective in the form of an affirmative proposition.

In the statement of objective (or hypothesis) it is essential to be specific as to the attribute being observed, the subjects on whom the observation will be made, and the questions of time and conditions. A statement of the objective or hypothesis for each of the four observational methods entitled in the figure Organization of Observations might be:

> Objective: To determine the number of books (i.e. more than 80 pages) in the hundred Dewey classes borrowed by male young adults between 18-20 years of age.
>
> Objective: To determine the size, composition and frequency of loans of library materials in relation to stated reading interests of young adults between 18-20 years of age.
>
> Hypothesis: Students with high academic performances in a particular school year will be shown to have differences in smoking habits during the year when compared with students with low performance.
>
> Hypothesis: The application of a longer loan period (4 weeks instead of 2) with automatic renewal will result in reading patterns that show less subject dispersion.

In the identification and selection of subjects, the observer (researcher) must describe in detail not only who (or what) are the subjects, but also how access to the subjects will be obtained, as well as how a selection of the specific ones utilized will be made from the larger pool of potential subjects. The observer must be acquainted with the problems of sampling error, bias in sampling, selection, techniques of randomization. The investigator must describe how he has given cognizance to the minimization of error and bias. The investigator must be aware of how to generalize to a larger population on the basis of his observations. If a control group is to be employed, the criteria for establishing the similarity or comparability of the individuals within the two groups must be described.

In describing his procedures for making and recording observations in detail, the observer (researcher) thereby develops a checklist of elements (i.e., the measurement scales) and the record forms required. He is reminded of the various specific observations to be made and hence of the criteria required to establish the categories in the scales to be recorded. It is customary to include, or append to the research protocol a specimen of the form on which observations will be recorded.

The record form facilitates the task of identifying areas of possible ambiguity. It is also helpful in establishing criteria and cleancut definitions to plan how the data collected will be tabulated on IBM or other devices. The establishment of a code requires an identification of specific data, definition of terms used, and criteria for each but of information to be entered on such a card.

The organization and presentation of findings is probably the most difficult task in developing a study because the observer (researcher) is required to anticipate the way in which data will be organized. If this task is performed well, it will be one of the most useful processes involved. This is why the formal statement of research design is eventually necessary. By anticipating tabular presentation, the observer (researcher) determines whether the data thought to be necessary for measuring the hypotheses will actually result from the investigation. When the proposed organization of data is viewed, the investigator can determine whether the objectives of the investigation have been met. A third benefit is the opportunity to note the probable difficulties in statistical analysis (from the projected sample size to the fragmentation of data compilation). The investigator utilizes the organization of data as a check on his sample and his procedures (for making and recording observations.)

The analysis of data and conclusions to be drawn, must of necessity await actual collection. However, alternative possibilities can be anticipated and examined in the probable analysis of data and conclusions. Often the investigator finds that he omitted collecting pertinent data while the opportunity was present. An attempt should be made to anticipate findings and increase the efficiency of the data collection. A pilot run on a very small group helps in anticipating problems. However, any pilot run should be considered as a part of research preparation rather than as part of the study itself.

## PRIMER ON THEORY

It may seem elementary to present a discussion of the microelements of logic such as observations, judgements and propositions. But if research design considers anything, it is certainly composed of these elements in patterns which make sense to the communicator associated with information, library and media science. The composition of the observations, judgements and propositions of necessity by the nature of the field are related to the dependent and independent variables with which the professional communicator is concerned.

Research is an extension of creative and critical thinking into logical designs. The purpose of research design is to assemble enough evidence so that conclusions can be drawn. The conclusions result from a rigorous argument whose logic is compelling enough to remove reasonable doubt about the judgements made. Research design adds another dimension of logical analysis to the usual methods of critical thinking in the problem-solving process.

In perusing the literature of information, library and media science (1), it is difficult to find any discussion about explicit theories of communication. The epistemological nature of man apparently is presumed to rest upon whatever humanistic school to which the individual member of the profession subscribes. In library science, the theory of communication appears to remain largely confined to that of the sociological and descriptive-linear model.

Since the humanistic schools of epistemology have spent almost as much time as the social sciences in their own specialized areas of inquiry rather than cooperating to understand man, communications science has based its approach on the cybernetic functions of the adaptive control organism with considerable success. Of course, the great systems of human thought have contained some concept of what man is. However for the purposes of communication science, man is the sum of those ideas and observations which have been submitted to the test of scientific evidence. Man is defined as the sum of systematically collected and empirically measured bits and pieces of scientific evidence.

Information science, on the other hand, appears to be reaching for a stochastic theory of information automata. The stochastic system has considered two major components. One component resembles a turing or information machine which is a set of states undergoing transformation. The operation of the machine rests upon information handled in bits. It can be defined recursively and described in two-value logic. The second component is a network information machine where a set of points are joined together by specified relationships.

Information science studies the transformation of human knowledge in order to provide an indexed space so that patrons may locate information "surprises" relevant to the nature of their inquires. Transformations may be isomorphic but retrieval is based largely upon descriptor sets that are presumed to be homomorphic transformations of the knowledge sought. Information science, however, has not so far been able to overcome the charge of descriptive labeling which Fairthorne (2) has leveled at its indexing practices.

Media science is concerned with the organization of nonverbal message space and with methods by which audiovisual stimuli can move initial perceptual confusion into affective and cognitive value. In the common daily life of men, nonverbal message space accompanied by some spoken language is the prevalent "library" of the people. Organized nonverbal message space has high instrumental value for promoting creative and critical thinking especially in probing for a verbal order in the perceived entropic happenings within which the adaptive control organism is involved.

In addition to social urgency there is an intellectual urgency for the development of a validated and verified theory of communication for library and information science. It may be that the profession has no function in communication other than the shibboleth, "The right book for the right reader at the right time." But an anomaly exists between what leaders have prescribed and what the profession as a whole will accept. There is nothing particularly strange in such a situation. But the time has come when leaders need a convincing, theoretical justification for their prescriptions.

The cybernetic model of communication has immediate implications for the library and information profession in two areas: counseling versus retrieval, and in communications producing situations versus message design. Both areas are significant: the first for an understanding of the difference between communication and information, and secondly, to acquire and organize knowledge spaces where information surprises can occur that are relevant to people in community contexts. Information is the input stimulus that "surprises", or modulates the perceptual apparatus of an adaptive control organism or mechanism, the perceived information is processed (interpreted) by the interpreter or governor within the adaptive control organism and some reaction occurs. When evident, the reaction appears as behavioral output which is fed back to the perceptual apparatus of the organism and/or to the parameter (s) which stimulated the original imputed information surprise.

Communication theory includes the conditions, materials and forms of intellectual examination, within and upon which meaning can be engendered in others and eventually communicative activity. Subject matter is data organized by some theoretical constructs as well as logical rules for observing and organizing perceptions by some theoretical constructs. Communication makes a metaknowledge possible, a knowledge of the intellectual instruments which produce knowledge. Thinking is not only directed towards the logic of concept relationships and of judgements about the experienced world but also towards the models of thinking and the resulting judgements about experiences made possible by the models.

Communication is concerned with the intelligence by which ideas are produced, tested and recreated. Communication attempts to develop in individuals the ability to think and to create new ideas and modes of engendering meaning in others. Abstraction is needed in communication in order to explore, describe, explain and invent. The level of abstraction should be appropriate to the specifics of professional experience, as well as the processes of communications programing.

Communication explores the ways in which description, explanation, reasoning and invention result in professional thinking, that is how models of thought are produced, tested, extended and otherwise affected. Certain procedures are common to all subject disciplines--exploring, describing, explaining, reasoning, and inventing. Mathematical (deducing conclusions from axioms), scientific (combined deductive - inductive) and philosophical (understanding of entailment) modes differ from the communication mode (models of thinking in the process of reasoning, judgements which results, and the form, range and quality of the communications producing situation made possible by the models.)

The general model proposed elsewhere (3) suggests that in the experience of any human being a continuum exists wherein all the skills of intrapersonal, interpersonal and audience communication occur. This continuum is not linear but cyclical and moves from one dimension to the other with flexibility and ease, at least in the mature person. Based upon the intrapersonal communication skills of reading, viewing and listening, the individual is constantly involved in cognitive development, cognitive flexibility and discourse message design.

The output of the cybernetic model of perception-interpretation is in the form of such grammatical units as questions, observations, or judgements. These grammatical units are formulated to serve as the basic elements of thought as well as of communicative discourse. Also in thinking, the individual is involved with the validity and reliability of his observations and judgements as well as the congruence between his questions and his experiences. Both inductive and deductive logic are used in his efforts to make himself appear as a consistent human being to his fellow men. Consistency is a function of the reliability of observations and the validity of judgements.

Reliability comes from the way in which a person makes observations about reality (his experience context) and depends upon the demonstrated congruence between his statements or observations and that reality which is experienced by himself and others. Of course the individual is free to pioneer and develop new methods of observation; but convention is a forceful element in determining the methods of observation that can scarcely be ignored.

Truth is the correspondence between the state of events indicated in a sentence and events as they actually exist in the real world. Because measurement scales are often difficult to develop the professional person all too often tends to interpret and evaluate what he sees rather than merely to report it. As Meredith would maintain, there is a referential element in all discourse units as distinct from the connotative or judgemental factors involved (4). Observations are composed of words which have denotative meanings (objects in the physical world) or structural meanings (relationships in the physical world).

Validity as distinct from truth is a product of inference both inductive and deductive that the individual uses in developing an argument (a discourse message). Conventions of logic of necessity play a more powerful role in judgements than in observations. Inductive and deductive inferences have a long tradition. The rules of inference have become solidified by the great weight of the discourse units which libraries house and by librarians whose "social" responsibilities are limited to the collection and organization of everything placed in the public domain. Obviously for the uninitiated, libraries (and sometimes librarians) are frightfully complex and disinterested places which are avoided by the majority of people.

In the history of human civilization logic did not develop until writing had occured and sufficient discourse units had been written making codification possible such as dictionaries and encyclopedias. In order to create sequences of thought and to analyze them, conventions of expression appeared which eventually developed into formal logic. Great libraries such as the one at Alexandria and elsewhere became laboratories where the first principles of bibliographic epistemology were tested and developed.

However, a few courageous thinkers do not let the judgements of the past which libraries aid and abet get them down. They use collections rather than the encyclopedia as a source of relationships in order to improve upon the historically determined and almost solidified corpus of judgements (discourse units). Their thought creates new discourse units more in tune with the times. In so doing, their judgements are enlivened by the evolving value systems of individuals, reference groups, social roles, and cultural movements and determinants.

Judgements are sentences that emphasize connotative meaning. They incorporate words which have sign-to-object-to-person meaning and relationships. Judgements tell a great deal more about social reality that they do of physical reality. While judgements are not empirically verifiable, their denotative elements can be tested factually. Judgements are believed or not believed, and in the final analysis can only be tested to the extent that they are acceptable to other people, therefore the significance of social acceptability, logical inference, and such principles of codification as exemplified by Merrill's Code for Classifiers (5). Documents are classed for their usefulness not the validity and truth of their contents.

Judgements are statements of value and exhibit the use of "good-bad" words which are an indication of intention i.e., what the source is telling us about himself, not about any particular object in physical reality. Since human beings are judgement-prone, inferential content analysis is often used to analyze discourse units for the intentions and purposes of the message sender. The librarian's methods of selection are an operational way of anticipating the impact of a particular discourse unit upon a particular audience grouping. Market analysis could carry the method one step further in order to anticipate and indeed "create" demands in particular audience groupings.

Judgements cannot be avoided in human discourse. In fact, there is a judgmental dimension in all observation, because complete objectivity is impossible, at least for human beings. Judgements often include persuasion whose purpose is to gain acceptance of new ideas, create changes in attitude, or to win votes for a new practice. The goals of judgements are persuasive, to affect behavior by making statements of value about the world.

Obviously perception must be selective if it is to be symbolized and reported at all. No human being is able to perceive all sense stimuli. Most are ignored, some few may be repressed, but only a fraction of stimuli are actively perceived, symbolized and interpreted or coded by utilizing the previously stored symbols. Judgements are unavoidable because physical reality is largely constructed rather than discovered by men. Structure and relationship are largely imposed by men upon the cosmos and remain so for periods of time as long as no incongruence with reality has been demonstrated.

Measurement is essential to ensure the reliability of observations so that someone else or even the same person will obtain the same observations in a similar situation. Measurement employs scales. These scales are called nominal, ordinal, interval and ratio in the order of increasing reliability which the measurements give. The more rigorous the reliability the greater is credence given the information-modulation being measured. The steps in making observations to ensure reliability include the following:

> Has there been a sufficient number of observations, for example at least 30 subjects in a sample? Is the evidence representative? Is the sample representative or might it be biased?

> Is there conflicting evidence? Do the same measuring scales yield different observations under similar circumstances?
>
> Have generalizations to causality been made upon inductive inferences when inferences to relationship alone are valid? That two things are related does not indicate cause-effect. Both may have been caused by an as yet unknown cause.

Inference is an essential ingredient of judgement construction. Inference may be inductive or deductive. Inductive inference is essentially based on the null hypothesis sometimes considered to be the foundation of "scientific" reasoning. Deductive inference has for centuries used the syllogism and has been replaced only in recent years by mathmatical reasoning and symbolic logic.

Inductive inferences are predictions about unknown events based on observations and the reliability of the measuring scale employed. Under the null hypothesis the inference of no structural relationship can be rejected. On the other hand it is not completely safe to accept the assumption that a structural relationship does exist. The resulting judgement is always tentative as the researcher considers the effect of types I and II errors. The judgement may have to be rejected when more observations are measured. In any event, the steps in inductive inference include the following:

> Develop a probability model, usually a model involving the null hypothesis which states that events X and Y are independent, and that there is no structural relationship between tehm.
>
> Define the cutting point between the rare or abnormal event and the normal event. In terms of the probability of occurence under the null hypothesis, cutting points are established at the .01 and .05 confidence levels.
>
> Use a test of statistical inference to decide whether the events observed fall on the rare side or the normal side of the cutting point.
>
> If the events fall on the rare side, reject the null hypothesis (because no one believes in rare events) and make an inference about how X and Y are related.
>
> If the events fall on the "normal" side, accept the null hypothesis and do not expect that inferences can be made about any relationships.
>
> However the researcher may have been in error, whether or not an inference was made. He may have done so when he should not have. He might have refused to make one when he should have done so.

Deductive inferences are logical arguments based upon "position papers" (or sentences), as opposed to the null hypothesis generation of principles in induction. These position paper statements are acceptable to two or more people who have extablished them for the purposes of conducting some argument. These seminal sentences (or position papers) entail (imply) other sentences. Over the centuries rather elaborate rules have been developed whereby entailments can be tested for validity. Symbolic logic has reduced these rules to a few logical operators for testing the structural rigor of the relationships being argued.

> Logical validity is independent of the concept of truth. Through deduction, sentences can be generated that have been implied by other sentences.
>
> Argument is not related to denotation. Argument does not necessarily need to accept any observations about physical reality. Argument manipulates judgements whose arrangement is internally consistent, that is valid logically.
>
> A sentence can be deduced from the relation between two other sentences, but if the two other sentences are true, then the deduced sentence is true. Consequently, statements should be tested and analyzed before making or accepting them.
>
> In presenting the conclusions of arguments, the entire argument should be reconstructed so as to be sure of its validity. Predict the extent to which others can reconstruct the argument for themselves or will accept the conclusion without reconstructing the argument.

Research is not emphasized for its own sake nor for the social approval accorded the scientific method, but because the conceptual tools used define the terms for what the researcher believes is going on "out there" in the world of actuality. The object of research and theorizing is to regulate the naming processes and hence the veracity of people's concepts. Everything spoken is not necessarily true. Messages and communication systems are not always necessarily related to reality. What people say quite often depends upon the channels being used, their own needs for status and identification, or that of their listeners.

It would appear as is meaning as an entity or as a message can be completely isolated from originator and receiver. This is particularly critical in library science where materials are dealt with as discrete entities and which of themselves are thought to contain meaning. The power of the written word still predominates among librarians almost to the exclusion of audio and visual "words." Apparently a single linear reading of a passage in a book is considered sufficient to establish understanding. It appears almost as if meaning could leap out from the printed page, and indeed from an information retrieval system, while the eyes continue to negotiate discrete word symbols sequentially without benefit of iconic symbol.

Reading is so conceptual as is surely information retrieval, and apparently tied to a logically related or aristotelian concept of meaning. Certainly it is in marked contrast to a perceptual theory where meaning is expected to occur as a result of interaction among all the senses of the recipient stimulated by his experiences based on viewing, listening and reading. There is nothing particularly wrong with the logical theory of learning, nor for that matter with the single reading, except that neither accounts for the full complexity of the human individual. The tragedy of the situation lies in the fact that librarians subscribe to the autonomy and dignity of the individual on one hand, but on the other confront him with service programs that are based on an oversimplified and now more or less archaic theory of communication.

Persuasion, or design of appropriate (meaningful) messages, employs not only the phenonmenon of definition and the associated phenomenom of redundancy but also several methods of interface with the receiver. Method is related to agency policy by which the agency representative (communicator) establishes a relationship, based upon purposes, with the constituents (receivers) which the agency is attempting to reach. These are general communicative methods designed to motivate participation and learning as distinct from the techniques (the communication or learning to be diffused) transcends both the method of interface and the channel (procedure) to be used.

Redundancy is a fact of man's existence. All human experience, societal, environmental and linguistic is redundant which is the complement of entropy or uncertainty. Redundancy helps denotative meanings by generating other meanings through the formal relationships which exist between them and unknown words as well as phenomena of experience. In this way hunches (insights) occur as to the denotative meanings of unknown words or unnamed experience. Not only does redundancy have great value for research, especially for the indentification of what to measure, but also without some redundancy perhaps no communications situation could ever be effective. Of course the problem is to hypothesize about intentions when a sufficient amount of redundancy has occured as for example when counselors and discussion leaders do it through probings or soundings as they are sometimes called.

Persuasion is the functional and epistemological outcome of observations and judgements. Depending upon the reliability of measurement and the validity of inference, the observations and judgements of message design should communicate the "reality" of the sender to the receiver.

Such an expectation may remain an ideal and often remote so far as human beings are concerned because the message modulates the very channel (human perception) over which it is being transmitted.

A study of definition is of considerable value to librarians both in message design (communication) and in research design. The function of research is to find new definitions (meanings) based on reliable measurement scales while that of message (program) design is to formulate definitions cooperatively with the audience group so that both sender and receiver understand one another, i.e., formulate common meanings. Therefore, as much as possible, cooperative involvement between sender and receiver is fundamental to communicative effectiveness.

Denotative meaning is concerned with the physical world and the existence of objects. It describes and "establishes" relationships (i.e. mentally) between words and the objects that exist in the physical world. Denotative definition is the attempt to name something in the physical world and to describe its existence. Denotative definition is a product of measurement and observation. It is operational and closely related to nonverbal behavior.

Such operational definitions are compiled in and provided by the handbook and manual where formulas and methods of procedure are defined, or lists of ingredients and sets of instructions are supplied. When the instructions are performed on the ingredients the individual gets what he set out to define. Of course the operational definition cannot give all the meaning of the term or thing which is being observed. The meaning has been reduced to that which applies in a given situation and in order to express a particular meaning.

Connotative meaning is person-oriented as opposed to the reality orientation of denotation. Connotative meaning is concerned with social reality. Connotative definitions do not report much about the world, but tell a good deal about the values, attitudes and judgements of the people making them. The use of terms such as good, desirable, beautiful make it almost impossible for two people ever to agree about connotative definitions. Obviously measurement is not precise, based as it is upon the nominal scale. However Guttman scales (6) and the semantic differential by Osgood (7) are attempts to provide operational definitions for connotative meaning.

It should not be construed from the preceding discussion however that formal meaning and definition do not have any value. On the contrary, unless there is some level of theory development and model building, i.e., analogous to the structural and contextual (intuitive insight) elements of language, it is impossible to do any research at all. The scientific method as we know it is both inductive and deductive. Even the null hypothesis makes some primitive assumptions about the structure of the cosmos. In any event, Berlo (8) has put the matter succinctly:

> In summary, the genus et differentia and other formal definitions are very useful when we are attempting to classify terms as members of classes, and separate terms from other terms that are members of the same class. The criticisms made are not leveled at the definition form itself, but only at our sometimes unfortunate uses of it for purposes it is not adequate to serve. The formal definition is useful as a shorthand technique to improve the efficiency of our communication and as a tool of logical analysis which enables us to manipulate terms and propositions carefully and precisely. It can be used when we are talking denotatively, but only when we are certain that there are physical events corresponding to our terms, and when we are not concerned about securing great precision of meaning.

Formal definition is the traditional method of placing a term in its proper class (genus) and then distinguishing it from all other terms in that class (differentia). Definition by class membership and definition by assigning properties are equivalent methods. However there are three limitations to formal definition: static rather than dynamic; universal rather than particular applications; relationships to physical world tend to be ignored.

Formal definition is static rather than dynamic. Terms are placed in classes and sub-classes and the meaning must be fixed and not relative to space, time, or personal viewpoint. Things either are something or they are not. As such they are based on a two-valued orientation. Of course two-value logic is the antithesis of cognitive flexibility and most people soon learn that such definitions cannot be maintained in the face of situational imperatives.

Formal definitions tend to be universal rather than have particular applications. Such definitions are an essential element in classification systems and typically use a form of "to be" in their formulation. Definitions of this type tend towards one view of the world and are the antithesis of perceptual sensitivity and transactional psychology where observations are determined partly by what is perceivable and partly by who is perceiving.

In formal definition, relationships extended to the physical world tend to be ignored. Classificatory definitions only implicitly relate signs to objects in the physical world. They are mainly concerned with formal reality. Like mathematics, the terms can be defined with great precision but they do not necessarily refer to anything. By using such definitions, e.g., democracy, an individual may be talking about something that has little if any relevance to life as lived in actual communities. Also in the counseling situation such definitions may get in the way of real needs. The patron may grasp onto ideas and psychoses that bear little relevance to his experiences. Therefore "talking-out" the problem is absolutely essential to achieve congruence and avoid self-deception and even psychoses. Librarians, for example, are desperately in need of training in definitions other than the formal ones to which they as professional people are especially prone.

The processes of definition are the means employed by a communications source to motivate people (receivers) to become involved in reciprocal communication. Definition is at the heart of model building which Bleth (9) has pointed out is so essential to creative and critical thinking as well as problem solving. In order that receivers will engage in the learning (communicative) enterprise for which the message is being designed, certain factors must be realized in constructing definitions which Berlo has summarized:

> The process of definition is essential when we are confronted with confusion or uncertainty about the meaning intended through our use of a term.
>
> Definitions do not provide us with knowledge. They are not statements of fact that can be said to be true or false, right or wrong, good or bad.
>
> The criterion of a "good" definition is its usefulness in specifying intended meaning -- nothing more.
>
> Deciding how to construct a definition depends on:
> - Intentions of the source.
> - Needs of the receiver -- why he is confused, and what he is confused about.
> - Existing meanings held by the receiver that can be used to achieve the purpose of the sender.
>
> There are two dimentions of purpose: structural or denotative meaning. Each has its own appropriate set of definition procedures, which cannot be used interchangeably. Both formal (structural, relational) and operational (denotative) definitions are necessary. Use is determined by purpose. Neither is sufficient alone.
>
> Formal definitions are a kind of shorthand, which permit the substitution of one term for another and the analysis of terms logically. The genus et differentia definition is the major formal definition.
>
> Operational definitions (i.e. measurement scales) are attempts to relate human concepts and terms to objects that exist in the physical world. They help us determine what it is that we refer to, or indeed whether we actually refer to anything at all.

Definitions do not give us knowledge. They enable us to classify and relate to knowledge previously obtained (formal) and to determine whether the potential for knowledge exists within the terms being used to talk about the world (operational).

## References

[1]Research in Education. Washington: Educational Resources Information center, 1966-

[2]Robert A. Fairthcrue. Towards Information Retrieval. London: Butterworth, 1961.

[3]Patrick R. Penland. Communication for Librarians. Bookstore, University of Pittsburgh, 1971.

[4]Patrick Meredith, "Documents, Programs and Topics." in Barbara Kyle ed. Focus on Information and Communication. London: ASLIB, 1965.

[5]William Merrill. Code for Classifiers. Chicago: American Library Association, 1939.

[6]Louis Guttman. "Cornell Technique for Scale and Intensity Analysis" in C.W. Churchman ed. Measurement of Consumer Interest. Philadelphia: University of Pennsylvania Pr., 1947.

[7]Charles E. Osgood. Measurement of Meaning. Urbana, Illinois: University of Illinois Pr., 1957.

[8]David Berlo. Process of Communication. N.Y.: Holt, Rinehart and Winston, 1960.

[9]Marc Belth. Education as a Discipline. Boston: Allyn and Bacon, 1965.

## Additional Readings

Marc Belth, "Diogenes Ascending." Advisory Counseling for Librarians. Bookstore, University of Pittsburgh, 1969.

Ralph Borsodi. The Definition of Definition. Boston, Porter Sargent Publisher, 1967.

Noam Chomsky. Language and the Mind. Harcourt, 1968.

Roderick A. Jacobs & Peter Rosenbaum. English Transformational Grammer. Waltham, Mass., Blaisdell Pub. Co., 1968.

Alfred Korzybski. Science and Sanity, an Introduction to Non-Aristotelian Systems and General Semantics. 4 ed. International Non-Aristotelian Library, 1958.

Susanne K. Langer. An Introduction to Symbolic Logic. Dover, 1967.

Daniel Lerner. Evidence and Inference. Glencoe, Free Press, 1959.

C.D. Needham. Organizing Knowledge in Libraries. Deutsch, 1964.

W. Van O. Quine. Word and Object. M.I.T. Press, 1960. Chap. 2.

Peter H. Salus. On Language. Holt, 1969.

## MEASUREMENT DESIGN

Measurement is a technique that underlies every judgement. Ordinarily a person is judged on first contact, partly in self-defense. In that sense the examination is tempered by personal fears and needs which become the determiners of objectives in the appraisal. In general terms, effective measurement answers questions such as the following: Where are you heading? Are you moving fast enough? What causes the movement and can it be improved? Does your course of action need to be altered? A few general characteristics which underlie the concept of measurement include the following:

> Measurement is always comparison, either qualitative or quantitative. It is therefore essentially the science of classification.
>
> Measurement is always a sampling which is determined by a philosophy or purpose.
>
> Measurement is the collection of facts and therefore can only be based on figures when the said figures have meaning.

Measurement is closely related to the fundamental question of the nature of objectives in research design. It is concerned with specifying the objective and of measuring that objective rather than with the specificity and accuracy of standardized testing. The researcher should avoid devising scales simply for the message content (independent variable) but focus upon the objectives (behavioral outcomes). Behavioral outcomes are being measured in order to infer about changes in the subject's apprehension of knowledges.

In relation to philosophy or purpose, social science measurement today stresses social values and adjustment with the idea of pattern in personal development. It is essentially derived from Gestalt psychology and recognizes the problem of measuring intangibles, but is applied on the basis that many of the important things, however difficult, require measurement. The present philosophy is somewhat of a mixture of the scientific or standardized testing and of the somewhat latter period of wholism or evaluation.

The present period tends to be non-elementaristic and to apply itself to the services or personal concept rather than the old idea of separating the normal from the atypical. More attention is paid to affective than cognizant factors. There is less faith in the idea that order exists as a single entity or that correlation is necessarily causal. More concern is expressed over organization and combination of measurements in specific cases than there is in the derivation of general laws. There has been an increased recognition of philosophic determination and involvement.

The contemporary period might then be characterized as leaning toward molarity and usefulness as opposed to the specificity and accuracy of standardized testing. This is an age of doubt as far as measurement goes. Change and controversy fill the air. As a result, it is necessary that professional personnel understand and be able to defend their own services to patrons. Much of what is done has to do with the surprise value of information as well as the communicating of information, skills and sometimes attitudes. Basically a measurement program is needed which would give an answer to how much is learned from either a particular message or a whole group or set of messages. One needs a method to determine how changed or how different people are as a result of being exposed to one of several programs in communication.

Order is characteristic of good measurement. The implications of measurement are such that logic and order are ordinarily more characteristic of it than of other phases of communication. It is hardly conceivable that any program of measurement, testing and appraisal could be adequately conducted if in that program there were not a consistent and orderly pattern or philosophy and procedure. The problems of format are essentially the problems of order in the mechanics. The major contributions of good format are toward accuracy and decrease of expense. But good format also contributes toward the morale of respondents and staff.

No one questions the fact that data is needed in order to test hypotheses in research design. It is almost tautological to say that data are necessary for research. However, it is not at all self-evident what data are, nor where these mini-facts come from and their relation to the hypotheses. In actual fact, data are the measurements or scores which have been taken from the outcomes of subjects submitted to an independent variable in the research process. The outcomes, whether books from a technical services assembly line or changes in human beings are always behavioral in the sense that they can be observed by some instrumentation.

The changes which result from administering the independent variable is the progress which the subjects have made towards the accomplishment of the objectives, that is the behavioral outcomes or more formally the research hypotheses. The objectives indicate what the researcher "believes" to be of value for the subjects to change to, or "achieve" as a result of being submitted to the independent variable. The relationship which exists between the statement of the objectives and the evaluation of achievement is a strong one. Not only is the statement of objectives essential to the evaluation, but it may be argued that, conversely, some attempt at evaluation is essential to the existence of the objectives.

The evaluation of achievement requires the ability to identify those characteristics which the subject has achieved and which distinguish him from the subject who has not achieved that objective. The researcher must be able to look at two subjects and say that one subject has achieved more in relation to a particular objective than the other. Such distinctions are sometimes difficult to make. If, however, it is impossible to distinguish between those who have achieved the objective and those who have not, it would seem difficult to defend the existence of the objective in any real sense. If the researcher does not know enough about a given objective to be able to evaluate its achievement, he will not know enough about it in order to conduct useful research.

A close parallel exists between teaching and research. This similarity is explicated in an attempt to clarify some of the "mystery" surrounding research development. All communicators may not have had the experience of teaching and testing. But they do have considerable experience in learning and in taking examinations. As students, they have frequently submitted themselves to the independent variable of the teacher and his message. Subsequently, these students were measured by the examination method. The examination is the teacher's method of collecting data about changes in the students towards some content (independent variable). It is assumed that the students (subjects self-selected from a population) will rank themselves from low to high in a normal distribution in relation to course content.

In education, for example, one type of objective which presents considerable difficulty is that of appreciation. Appreciation may be of art, music, or literature, or the contribution to modern life of science or history. Consider the area of literature and especially the study of Shakespeare. To develop an appreciation of Shakespeare in students is certainly not hard to justify. The difficulty however arises when the attempt is made to identify characteristics of the individual who appreciates Shakespeare. It is necessary to identify such characteristics in order to distinguish him from the individual who does not have such an appreciation.

It may be possible to reason that the person who appreciates Shakespeare is more likely to be able to identify plays, describe characters, quote passages, or identify quotations. If this is "appreciation" then the teaching of appreciation becomes the teaching of these behaviors. However, one soon realizes that while the individual may have learned to perform these tasks he at the same time may have developed an abhorrence of everything Shakespearean. The problem lies in accepting these behaviors as a definition of appreciation as well as in the difficulty of identifying the characteristics of the person who "appreciates." If it is impossible to do so, then it is also impossible to teach appreciation.

It is not the intent here to say that the development of appreciation can neither be taught nor researched. However, any proposal for the teaching of appreciation as well as any proposal for its research and evaluation must begin with an identification of the characteristics of the individual who has appreciation as distinct from the person who does not. Unless this can be done for any subject whatsoever that is to be researched, no research can be undertaken. Fundamentally, the problem results from the question of ultimate versus immediate objectives.

One of the principal problems to the measurement of change or achievement comes from the fact that the initial concern may be with ultimate objectives. This means that the knowledges, understandings, habits, and skills which the subjects are supposed to have cannot be obtained within the research situation and time span. The study of communications in information, library and media science may serve as a case in point. The dichotomy between what librarians for example say they do and what they actually do has reached the proportions of a mental dislocation in the profession. On the other hand, the subjects in a research design may be expected to change to an objective which is so encompassing or so developmental that it would take a long period of time to achieve. If however such a research study must be conducted, the researcher may have to use a genetic method and patiently wait until such time as data accrue from the measurement instruments.

Of course not all objectives are so grandly ultimate. In many cases the knowledges, attitudes and skills with which the observer is concerned can be made susceptible to measurement scales. The researcher is forced to accept or to identify immediate objectives. In so doing, those immediate objectives are identified and selected which will lead most directly and efficiently toward the realization of ultimate objectives, ie. create the theoretical constructs necessary to the support of communication in information, library and media science.

Sometimes an immediate objective may be expressed in terms identical to the ultimate objective. Objectives in the skill's area can more readily be of this type. Immediate objectives should always be selected with theories or ultimate objectives in mind, and subjected to critical review in the light of those theories. Too frequently this relationship is ignored. All too often the immediate objective of the impatient researcher is simply for factual data. Because of this undisciplined orientation to research design, the field survey for example is all too popular and as a result does not realize the potential significance it might have.

Another problem in the approach to research objectives is that measurement design depends upon the identification of the behavior of subjects, ie. the variables of behavioral outcomes. It is so far impossible to measure knowledge or attitudes, only what knowledge does to behavior. It may be possible for a human subject to possess understanding of a concept and not display that understanding in behavior. If, however, the individual does not display the understanding in behavior which can be observed, no basis exists for assuming the existence of the understanding.

If the researcher were concerned with height as a characteristic, he could stand John beside Bill and make comparative statements, such as "John is tall," "Bill is short," or "John is taller than Bill." If he desires a more precise measurement of John's and Bill's heights, he can obtain a stick which has been marked off in inches (or any other unit) and make a comparison between the heights and the numbers on the stick. Such a measure however is based on a ratio scale which can have a zero point. A ratio scale is also an interval, ordinal and nominal scale. Current understanding of human behavior is so limited that such precise measurements are difficult. Consequently the nominal and ordinal scales are most often used to reflect human behavior.

With regard to the communications outcomes, however, comparisons cannot be made as precisely as height. One cannot look at John and tell to what extent he has developed the ability to read. In reality no one can look at John and tell whether he has learned to read at all. Nor indeed have any devices been developed which can be attached to John which will yield this kind of data. The only approach to the measurement of John's ability to read is through the effect of that ability upon John's behavior. The observer looks at John's behavior in a situation calling for reading ability. On the basis of this observation he draws an inference about the underlying ability to read.

The measurement of human characteristics by this approach is frequently used as a rationalization for the biases towards the social sciences which exist among unsophisticated specialists in information, library and media science. Measurement by effect does not, in and of itself, lead to inaccuracy. Few people, if any, would question the the thermometer as a measure of temperature, but the thermometer measures temperature only by effect. It measures the effect of temperature changes upon the expansion of an enclosed fluid or upon the expansion of metals in a specified arrangement. Furthermore, people generally accept the fact that the different characteristics of an electric current can be measured. Nor does it bother them that such measurements are possible at a time when there is no concensus as to the nature of electricity. In any event, electricity is not measured, only its behavioral effect.

The approach to the measurement of many human characteristics, including achievement, is exactly the same as the approach to the measurement of temperature or electricity. If the measures of human characteristics are less accurate, it is not because of the approach used, but rather because the knowledge of the effect of the characteristics upon human behavior is less certain. It should be emphasized that the difference in accuracy here is one of degree and not one of kind.

Measurement Criteria:

Objectives should be stated as the behavioral outcomes of subjects, rather than as research objectives. The research should concern himself with what the subject to be observed is able to do rather than with what the grand objectives of a profession may want him to do. A case in point, is the librarian's so-called "bill of rights for adults." These ultimate objectives may very well be socially relevant. However, the relationship has not been established between what patrons are supposed to do in libraries and what they actually do in coming to terms with personal information surprise. To say that the library's purpose is to provide the opportunity for creative reading is not an objective at all. It is rather a statement of part of a method, which the librarian intends to employ.

Since the researcher must rely on measurement through effect, he must be able to identify those behaviors which indicate the achievement of objectives. "To know," "to understand," "to appreciate" are not observable behaviors. The following, while stated as an objective must be stated in terms of behaviors which indicate the posession of the observable behaviors:

> "To be able to define each word on a list."
> "To pick synonyms and/or antonyms for each word on a list."
> "To use each word on a list correctly in a sentence."

Of course the most adequate evaluation of achievement would be based on the observation of those behaviors which indicate the accomplishment of ultimate objectives. For each ultimate objective there exists, theoretically at least, a series of behaviors which completely defines the achievement of the objective. This list includes all of the behaviors which a subject might be called upon to perform and which indicate the accomplishment, or lack of it, of the objective. Such behaviors define the particular ultimate objective and may be referred to as the criterion series of behaviors.

Any sample of behaviors from the criterion series has to be comparable from one individual to the next. All measurement derives its meaning through comparison. Comparison is not possible unless the behavior of one individual is the same as, or highly similar to, the behavior of others. For example, while content and factor analysis may be possible on the "subjective" statements of respondents, it is impossible to compare each respondent's style and treatment of the subject matter. In research terms this means that the sample from the universe of content which becomes the independent variable must be the same for all respondents.

The use of the total criterion series for evaluative purposes is limited by considerations of efficiency, comparability, economy, and expediency. No researcher can depend exclusively upon observing his subjects in natural situations. He is forced to create situations to which the subject can respond. In any event, the researcher should attempt to make the items in the experimental series identical with those in the criterion series. The closer the relationship between the experimental series and the criterion series, the better the measurement will be as well as the evaluation and subsequent extrapolation.

In other words, the items which constitute the independent variable should be closely related to the theory underlying the research design. The theory of course is expressed in the hypotheses. The independent variable which grows out of the hypotheses is the "course" which some subjects "take." These subjects however are not selfselected but consitute a random sample of a population. A second random sample of the same population is drawn but they are either not given the "course" or given the "course" by a different method. Each sample group of subjects is nevertheless given the same "test" on the content of the independent variable. Presumably there is a "significant" difference in the score obtained by the two groups.

The relationship between the independent variable and the test (examination) or measurement has considerable significance in research design. Presumably it is the function of the independent variable ("the course") to precipitate (communicate, teach) some range of behavioral outcomes. This range of behavioral outcomes ("Criterion series") is practically always greater in scope than any test can measure at least immediately. It is then necessary to have some guidelines to determine the behavioral outcomes in the criterion series which can be measured (tested). These guidelines in general have been considered to be the validity, reliability and usability of the measurement scales which go to make up the measurement instrument.

Validity is the first and most important criterion. Validity means that the researcher does measure what he sets out to measure. Error in validity creeps into measurement through clues and non-functioning elements. If the respondent can answer the question by a clue without knowing or having the required skill, then validity disappears. Error also occurs when a person who is likeable is considered to be able to do everything well. If he writes well, the observer may rate his content feedback higher than he otherwise would. This is sometimes called the halo effect.

Another error in validity may arise because the observer does not properly examine the universe of content. These content errors occur because a wide enough sample is not taken, or questions of varied difficulty are not used. Finally, statistical validity is when the test agrees with some judgment other than the researchers about how well the participant does. Ordinarily when validity is reported, it is on this basis. Validity then is a measure of whether the data that is wanted is indeed obtained.

Realiability is a measure of whether or not all the items as a whole in an instrument agrees with itself. In other words, if it is given again will it get the same results. If an instrument then is reliable, it is consistent. The same answer is obtained from the scales every time. This, however, is no proof of validity. It is possible to have a reliable test that is not valid. The scale may be consistent, but it is consistently wrong. But it is difficult to have a valid test that is not reliable. In other words, if the thing being measured is measured, then it will be measured the same all the time.

Reliability is a statistical concept. A longer test is more likely to be reliable. Obviously, an objective test is likely to be more reliable and as a result objectivity is one of the characteristics of a reliable test, because two competent users will get the same results. However, if the instrument is totally objective, the validity of it may be destroyed. There is, therefore, no completely objective measure possible. But when information is being obtained from respondents, objectivity may aid in securing reliability.

Usability may refer to the ease of instrument administration, the expensiveness of a test, or the possibility of interpreting and applying the results. For example, in centers where interest tests and personality tests are used, there may be extensive files of records which simply gather dust. There is no time and no personnel capable of interpreting them. Consequently the records of such testing become dead files. Such testing is obviously useless.

More specifically, these general criteria of validity, reliability and usability may be delineated in procedural steps for the individual scale items. The identical elements type of measurement scale is one in which the behaviors required of the respondent are identical in every respect with those behaviors being researched in the criterion series. Rather then wait for natural opportunities to occur in which to observe these behaviors, the researcher creates situations in which the respondent is more or less forced to display them. Many work sample or performance scales are of this type. However evaluation should be based upon the quality of the finished product. Its importance results only from that effect. Too frequently a disproportionate emphasis is placed on procedures in contrast to the finished product. Procedures become important only when the finished product needs improvement.

The related behavior scale is one in which the behaviors of the experimental series are related to, but need not necessarily be identical with the criterion series. Although the experimental behaviors may not be similar to the criterion behaviors, it should be possible to demonstrate a strong relationship between the two. Tests based on the analysis of complex processes into component parts and which measure separately the individual components are generally of the related behavior type. For example, a person's ability to save lives in swimming may be judged through his performance in a simulated situation. The observer, or another participant will feign drowning and the research subject will rescue the "victim." The behaviors required here are similar to the behaviors in the criterion series. The only difference is that no one is really threatened with drowning in the experimental situation. The simulation however is fundamental to the measurement of the behavior which will be evaluated.

Verbal and nonverbal behaviors can serve as the basis for measurement scales. In verbal behavior for example elements from the criterion series are presented to the respondent. He is asked to tell what he would do (or should do) in such situations. The situations are generally presented in verbal form whether written or oral. But the presentation scales may involve charts, diagrams, pictures or other materials. The respondent may be required to tell what he would do in each of the situations in his own words. Or he may select from several possible behaviors the one he considers best in the situation. The ability to perform on tests of the verbalized behavior type can be closely related to the ability to perform in the criterion series.

However, there are limitations to the use of the verbalized behavior scale. In some cases the ability to describe a behavior does not guarantee the ability to perform that behavior. The person who can describe in detail an acceptable procedure for driving a golf ball may not be able to hit the ball in actual practice. Conversely, the best golfer in the world may not be able to describe his approach. In addition, what the individual says he would do in a given situation and what he actually may do in that situation are not necessarily the same. In response to many items the respondent may give the socially acceptable answer, or a response thought to be favored by the researcher in order to please him.

In any event probably the most common scale is the type which attempts to determine how much a respondent knows about a given topic or subject. In applying this approach for example to the prediction of future voting behavior, the researcher may reason as follows: "Before an individual can vote, he must register. In order to register, he must know where to go to register. Therefore, if he cannot tell where to go to register, he probably will not vote." The researcher then asks the respondent: Where does one go to register to vote? However it does not follow that possession of the required knowledge will guarantee that any individual will indeed perform the desired behaviors. This is especially true in the case when our definition of "to know" is "to repeat factual data."

In addition to changes in knowledges and changes in skills that have to be tested for, perhaps the most subtle and most important are changes in attitudes. The initial level of attitude should be obtained just as it is in the case of knowledges. The changes in attitudes that are supposedly the result of a training program ought to be tested by administering some form of questionnaire before the training program and a similar form at the end. If possible, this also should be compared with scales administered at the same time to people who did not take the training program. Such people should be similar by having attitudes presumably the same initially as the trainees whose attitudes are not the same when the program ends.

The attitude scale is a variant of the skills-knowledge content of measurement design and may be considered as a major contribution of social science research. The study of attitudes is based on the assumption that attitudes serve as a link between knowledge and action (behavioral outcomes). The value of attitude study is for predictive power. Given strong attitudes, the individual will act in conformity with them if a situation occurs in which it becomes necessary to take action. Attitude surveys consequently reveal the climate of opinion at any particular moment on some particular issue. Surveys can be conducted periodically as in public opinion polls to reveal trends in public opinion. Attitude measurements can be made before and after an independent variable has been administered to a sample from a population. Or, the attitude instrument can be given before and after the sample has participated in an experience as for example an institute or training session.

Closed items tend to obtain fairly reliable or dependable answers. This is evident in a listed series of responses where one is correct for each item or in true-false questions. Matching items are also of this type where response categories on the left are to be matched with categories on the right. Regardless of who scores the response sheet, even a machine, the score will be consistent.

On the other hand, open-ended questions present more of a problem even though it is possible to achieve reliability. Usually, to get reliability the points in any response have to be listed that need agreement, such as: "So much credit will be given for mention of such and such a point. So much will be taken off for this or that kind of error." The response sheets have to be read and rated independently by the two or more people who are going to do the rating. The answers, not the differences have to be compared. It may be necessary to start all over again, until a satisfactory level of agreement is reached. Finally, the revised version is checked against the response sheets in order to obtain a satisfactory level of agreement between scores.

Other less formal observational techniques may be used as supplements to more structured instruments. Of course, a number of unobtrusive measures exist. One highly unstructured device is simply to have experts take a look at a program and give their judgments. Another way is to pay attention to, observe, and talk to people as they attend a program. This is probably the more usual way in which program chairmen test their effectiveness and determine the success or failure of communication sessions. After all, if people are really interested, they are going to pay attention. Their eyes will be glued to the speaker or the visual demonstration, or what have you. If they're bored, they'll be doing all sorts of other things. They'll be easily distracted; they'll pay attention to a lot of extraneous stimuli but pay little attention to the focus of the communication. At least, that is one way of measuring audience interest.

Another measure, of course, is simply to discuss the subject of the program with the audience afterward. Have them tell you what they thought about it, whether they thought it was worthwhile. Did they like it or didn't they like it? Did they find it interesting or were they bored? Many things come out just in ordinary, informal conversation. Such a procedure may not be quite as objective, as reliable, and as valid perhaps as the more formal methods, but it does help to evaluate what has been going on. At the least, it could be used to supplement other methods. Finally, perhaps the most important measure of the effectiveness of any service program will reflect itself in actual, subsequent and long-term behavior. In other words, if a program is designed to add information and change attitudes and behavior about child-rearing, the acid test is "How do these people behave with their children after receiving my program? How does their behavior now compare with such behavior as did not occur before the program?"

Sometimes the people who are innovative and forward looking are the very people who will come to a program. If these people are observed and compared with the people who did not come, it may be possible to identify the change which took place among the people who did come. It might well be that those people who did come would have changed anyway. However people who did not come are much more likely to be people who would not change anyway if they did come. One way of course to deal with the problems is to assign people to the programs at random. But this is impossible and the researcher has to resort to observation. The changes that do take place may be changes that could easily have resulted from things that took place in the programs. Whether the programs can take credit completely or not, may never be found out.

## Instrument Development:

In discussing the quality of a particular measuring device, the principal concern should be the extent to which the device serves the purpose for which it is used. All measurement involves approximation. Any "score" which is obtained from the application of a measuring device will contain some error. The score will be influenced to some extent by conditions, or factors, which have nothing to do with the characteristics being measured.

When an individual stands on a bathroom scale, the "score" that he reads (the number of pounds) will be determined only in part by his actual weight. The score will also be affected by the amount of clothing being worn and by the weight of the things in his pockets. The condition of the scale itself can influence the score. The adjustment may be inaccurate, or the scale may stick. The position of an individual on the scale as well as his angle of sight may have an effect. Even the individual's attitude about his weight (whether he considers himself overweight or underweight) may be reflected in the score he reads. The influence of any of these extraneous factors may be slight. But to an extent they can introduce errors in measurement programs.

The approach to improving the quality of the measurement of the individual's weight may be obvious. But it indicates that the researcher should attempt to identify as many of the sources of error as possible, and eliminate or minimize the effect of such error factors. For example, the errors associated with the individual's clothing can be eliminated by eliminating the clothing. The errors associated with temperature and the age and amount of use of the scale can be lessened by substituting a balance scale, since it can be demonstrated that these factors have little effect on such a scale. In any event, a similar approach to the improvement of quality can be applied to any measuring device, observational technique or paper and pencil test.

The validity of a test is the consistency and degree with which the instrument measures the trait(s) which it is used to measure. To be valid the instrument should be consistent in the scores that it gives. If it can be assumed that an individual has not changed regarding a trait measured, a valid instrument should give the same score for that individual on each of several administrations. This consistency of an instrument score is designated by the term reliability. The degree to which the instrument measures the trait which it is used to measure is the relevance or validity of that instrument. To have some degree of validity the instrument must have some degree of both relevance and reliability. If it is completely lacking in either relevance or reliability it can have no validity. To be perfectly valid (which no instrument can be since perfect measurement is impossible) an instrument would have to be both perfectly relevant and perfectly reliable.

Relevance of course is judged in terms of the use that is made of the instrument. An instrument may be highly relevant in one situation but less relevant, or completely irrelevant in another. An instrument designed to measure the knowledge and understanding of current events may be relevant today, but in ten years it will no longer be relevant.

The question then is whether or not an instrument is relevant to the purpose for which it is used, the group with which it is used, and the time at which it is used.

The reliability of an instrument on the other hand is the consistency with which that instrument measures whatever it does measure. A reliable instrument tends to give the same score for the same individual on each of several administrations. However, an instrument may be highly reliable even though the characteristic which it measures has nothing to do with the purpose for which it is being used. Reliability refers only to the consistency of the scores obtained for the same individuals on successive administrations of the measuring device.

The meaning of reliability used to describe a testing instrument is somewhat more limited than the meaning of the term in general application. When a person is reliable in keeping appointments it generally means that he is consistently on time. Consider, however, the individual who is never on time for an appointment but who always arrives fifteen minutes late. No one could argue that this person is punctual, but it can be pointed out that he is reliable. He can be relied upon to be fifteen minutes late. In this sense the individual who is always fifteen minutes late is as reliable as the individual who is always on time. His reliability depends upon the consistency of his behavior, not necessarily on its accuracy.

Considering the factors which affect a test score, we can note that it is the influence of temporary and chance factors which will cause a person's score to be different on one occasion from what his score would have been had the test been given on a different occasion. The reliability of a testing instrument, then, is inversely related to the influence of temporary and/or chance errors of measurement. A perfectly reliable instrument would be one on which the influence of all temporary and/or chance factors had been completely eliminated. Such a situation is possible only in theory. On the other hand, a perfectly relevant but completely unreliable instrument is impossible even in theory. To be completely unreliable all the variation in its test scores would have to be due to temporary and/or chance errors and none to content.

Any procedure for estimating the reliability of a measuring device will involve securing at least two scores from separate administrations of the instrument or from the administration of different forms of the same device. If one is interested in the reliability of a bathroom scale, several measures can be obtained of an individual's weight using this one measuring instrument. If the scale is highly reliable, any of several scores will all be the same. On the other hand if the indicated weight for one individual is 135 pounds on one weighing, 131 pounds on a second weighing, 142 on a third, and 128 on a fourth, it can be concluded that the scale is not very reliable.

The amount of agreement among the various scores is an indication of the degree of reliability of the scale. The estimate of the reliability could be improved by noting the agreement among several scores obtained for each of a number of individuals. It should be noted that in estimating the reliability of the scale, it is not necessary to know the real weight of the individuals being measured. To be reliable the scale does not necessarily have to give accurate measures of weight; it need only be consistent in the measures given. Of course the individual's real weight (his "true" score) has remained constant over all of the measures obtained.

This assumption is reasonable if the different weight scores are obtained within a short period of time. However, such an assumption may not hold if an attempt is made to obtain several scores from different administrations of the same test of educational achievement. In using an achievement instrument the individual will have changed simply through having taken the test. In this situation two different scores can be obtained for each of the individuals in a group by using some procedure for securing scores from equivalent forms of the test.

Having by some method acquired for each of a number of individuals two scores from the same instrument, an index can be obtained of the degree of relationship between the two sets of scores by computing the correlation coefficient. This coefficient is a measure of the reliability of the instrument since the more closely the two sets of scores are related the more reliable it is. If the instrument is reliable the individual with the highest score in one of the sets would also have the highest score in the other set. With a perfect relationship between the scores, one set of scores arranged in order from highest to lowest would correspond to scores in the other set.

Anyone of the three following approaches may be used to obtain two scores from different forms of the same instrument. The particular method employed in estimating the reliability of a given instrument will be determined by its nature and the situation in which it can be used. The interpretation of a reported reliability coefficient depends upon the method employed in the computation. The following approaches include the split-test, the test-retest and the equivalent forms as methods for estimating reliability.

The split-test method provides two scores for each individual from a single administration of the instrument. After the testing instrument is administered the items are divided into subgroups for scoring. Frequently a score is obtained for the odd membered items, and a separate score is obtained for the even numbered items. If the reliability coefficient is to be an accurate measure of the reliability of the instrument, it should be assumed to reflect the effect of the operation of temporary and/or chance errors of measurement. For example, if the performance of respondents is affected by the weather, weather introduces a temporary error into the test score and lowers the reliability of that score. For the effect of this error of measurement to be reflected in a lowering of the coefficient of reliability, the weather must be different when the two scores are obtained.

The principal limitation to the split-test method of estimating reliability arises from the fact that most sources of temporary errors in measurement remain constant over both sets of scores. An individual does not have a headache while answering the odd numbered items and lose it while answering the even numbered items. Reliability coefficients based on a comparison of odds-even score, then, tend to reflect the operation of chance factors but not the effect of other temporary errors of measurement. Some temporary factors (such as fatigue) will operate differently on the first half of the test than on the second half of the test. The effect of these factors will be reflected in the reliability coefficient when that coefficient is based on a comparison of scores in the first half with scores in the second half.

In order to offset limitations in the split-test method of obtaining reliability, it is possible to divide the items in the instrument in order to maximize the estimate. A better estimate of reliability can be obtained if the specific abilities measured by the two halves of the test are identical and if the items in both halves are of equal difficulty. If either of these conditions does not hold, some factors will vary in their influence on the subgroup scores and will be reflected, erroneously, in the reliability coefficient. In any event, neither condition holds for most achievement tests since items in the instrument are arranged by topic and in order of difficulty.

The test-retest method is the approach employed for example in estimating the reliability of a bathroom scale. In this method the same instrument is administered to the same individuals on two different occasions. If the occasions are selected properly most temporary and chance facotrs can be given an opportunity to vary in their influence on the two scores and thus be reflected in the reliability coefficient. However, this method is not always applicable especially in estimating the reliability of test of educational achievement. Frequently the individual has changed simply through having taken the test.

In this case any score for lasting and systematic factors (including possibly his "true" score) will vary from one testing to the next, and the reliability coefficient will be lower than it should be. On the other hand when items are repeated the individual has a strong tendency to repeat from memory the answers he gave on the first testing. This tendency is strong enough to cancel the effect of many temporary and chance factors which would otherwise operate to make his two responses differ. The repetition of previous answers will lead to an overestimation of the reliability coefficient.

Even in the case where the measurement does not result in a change in the individual, the selection of the appropriate time interval between testings presents a problem. If the test is repeated immediately, certain temporary factors will remain constant over testings, and the reliability coefficient will be inflated. If too much time elapses beween testings, the "true" score of the individual may change through learning, and the reliability may be underestimated.

The equivalent forms method eliminates most of the difficulties encountered in the other approaches. Care must still be taken to allow sufficient time between testings to permit the influence of temporary factors to vary and, on the other hand, to avoid allowing so much time to elapse that the lasting and systematic characteristics of the individual will change. Of course, it must also be possible to assume equal relevance for the two forms. This assumption is reasonable when the items included on the instrument are in reality a sample from the items which might possibly be included. The transformation has to be made in a logical manner and perhaps homomorphically.

Under certain conditions, statistical procedures can be applied in estimating the relevance of a test. Consider again the bathroom scale as a measure of weight. If the "true" weights of a number of objects (or individuals) are known, an estimate could be made of the relevance of the bathroom scale by comparing the weights it gives for these objects with the "true" weight. Of course it may never be possible to obtain a measure of the true weight of any objects since no device which can be used will ever be perfectly reliable. However, a measure can be obtained on some device which the researcher is willing to accept as a relevant measure of weight. For example, a balance scale may serve this purpose. If so, that scale provides a criterion against which the relevance of our bathroom scale can be computed.

Similarly the empirical relevance of any test can be estimated whenever some measure other than the test itself is accepted as a criterion of the trait in which the researcher is interested. This "outside criterion" is assumed to be a relevant measure of the trait. However, neither the criterion nor the test scores can be perfectly reliable. Through statistical procedures, the correlation coefficient between the criterion and the test scores can be computed and this coefficient "corrected" for the lack of perfect reliability in both of the sets of scores based upon regression formula. The "corrected" (i.e. adjusted) coefficient is an index of the relevance of the test.

Estimating empirical relevance depends upon the availability of an outside criterion of the characteristic measured by the instruments. Such a criterion may be difficult to obtain. If the concern for example is with educational achievement as defined by a particular librarian for specific respondents, no outside criterion exists against which the librarian can judge the relevance of his measures of achievement. Poor though it may be, the librarian's device will be the best measure available. In evaluating achievement he should use the most relevant measure obtainable. In any event, if an outside criterion is not available, an empirical estimation of relevance cannot be applied.

Estimating the logical relevance of an instrument involves an analysis of the behaviors required in the experimental situation and a comparison of these behaviors to those which define the accomplishment of the objectives for the research design. This, then, is a comparison of the test series to the criterion series of behaviors. To the extent that the behaviors of the test series are similar to, or may be assumed to be related to the behaviors in the criterion series, it may be assumed that the test is relevant.

In developing a relevant instrument, ultimate objectives must be kept in mind in order to select immediate objectives which will lead to ultimate objectives or theoretical constructs. The immediate objectives must be defined in terms of specific observable behaviors so that observed behaviors will lead to appropriate research judgements or conclusions. To the extent that these prescriptions are not met, the relevance of the research evaluation is threatened. Although difficult to attain, these standards should be met as closely as possible considering efficiency, economy, comparability and possibly expediency.

In considering the relevance of a given instrument, it is sometimes possible to establish the skills, knowledge, and attitudes required in the test performance as being closely related to the content and operations which have been presented in the independent variable. In such instances, the instrument is relevant to the immediate objectives of the research design although it is difficult to estimate its relevance to ultimate objectives. This limited type of logical relevance is referred to as the content relevance of the test. The test with a high degree of content relevance will possess a high degree of logical relevance only when it can be assumed that the achievement of the immediate objectives will lead toward the achievement of the ultimate objectives. An instrument has content relevance if it measures what has been presented in the independent variable, whereas it has logical relevance if it measures what should have been observed in the light of ultimate objectives.

## OBSERVATIONAL METHODS

A topic for research grows out of the commitment which the communicator has to information, library and media science. Deeply concerned over the orderly development of services, the communicator manipulates the infrastructure of information, library and media to meet the imperatives of rapid social change. As an information processing specialist, he serves society by working to reduce the time lag between the scientific creation of new knowledge and its use in sociocultural and political decision making.

The communicator also exercises his leadership role in society by articulating social concerns and interests with special relevance to those areas where knowledge does not presently exist. Once identified as lacunae in the corpus of knowledge, the communicator endeavors to motivate appropriate scientists to do research and create the needed but currently unavailable new knowledge. Admittedly surveillance in nature, no other profession can discharge this responsibility so effectively as the communicator backed up by the infrastructure of information, library and media science.

Finally, however, the communicator has a unique role to play as a researcher in his own right. As a communicator, the specialist in information, library and media science is mainly concerned about the surprise value of information in the lives of people. When data becomes kinetic, this information has the potential to effect change within individuals, groups and communities. Many years ago, Asheim (Library Trends, October 1957) summarized the matter succinctly: content analysis and audience research are the major research methods for the kinds of proglems faced by the communicator in information, library and media science.

Nothing has appeared to negate that prescription. In fact, the research lacunae of the intervening years and especially Goldhor's Measurement and Evaluation (University of Illinois, 1967) have made Asheim's statement more of an imperative than ever. While specialists in information, library and media science may get satisfaction out of being gestaltists as far as knowledge control (informatology) is concerned, the behavioral aspects of human endeavor receive but passing attention. It may be necessary to improve the liberal education of such infrastructure specialists in social science research. In any event, the human meaning (information surprise value) of the social science is evident in Berelson's Human Behavior (Harcourt, 1964).

This research imperative may be restated in the terminology of the preceding chapter. The ultimate objective is a comprehensive understanding of data made kinetic especially in the surprise value of information within individuals, groups or the community. The immediate objective is to identify the actual behaviors of individuals, groups and communities in using information. Once variant patterns have been observed, it may be possible to establish the criterion set which actually should underlie the communicator's purposes and policies.

In the daily professional life, many situations and problems occur which may serve as sources for research concern. Problems arise and explanations must be given for purposes and policies. Tragically enough, no one ever seems to be around who has any answers other than that is the way it was done before. This lack of reference to principles let alone a theory is irritating. It is like flouting the "Bill of Rights for Adults" before patrons and in the process coming to the realization that patrons are something less than enthusiastic about techniques to display the librarian's repetoire.

As a result, the communicator has a felt need to develop a better explanation than has been provided in the past or even by leaders in the field at the present moment. The dissatisfaction subsequently grows to the point where he begins to do some reading in the professional and related literature at first in an exploratory, heuristic way. A variety of points of view are expressed. Some may even coincide with his own thoughts. Indeed the problem may not be particularly new. However, there does not appear to be any concensus on an analysis, let alone solution to the problem.

But the discourse development in this literature is on the deductive and logically orderly level so necessary for validation. In order to look at the topic in an organized way it is necessary to read the research literature. As in other instances, the reading of research reports is essential to the processes of sustained and logical thought. This literature reading hopefully may help him to lay out the field, to identify major and minor issues, solutions and analyses. This reading helps the communicator to conceptualize and to develop cognitive flexibility about his felt need. This conceptual development is very important to the process of nominal scaling as well as of categorizing into discrete and ordinal sets the concepts in the field of the communicator's concern.

Possible solutions come to mind as reading continues, solutions that may be implemented by administrative decision-making, survey research or hopefully by research proposals. Continuing reflection is also necessary in order to synthesize the ideas which have been generated by the reading. Insights occur which have not been discussed previously and these insights may become strong enough to support an hypothesis which can be submitted to research design. In effect one is willing to consider making a judgement and risk his integrity upon that statement.

Once this occurs, then the researcher goes back to the literature and reads it from that point of view. His previous tentative statement becomes a tentative hypothesis while the analysis of previous research studies continues. The literature being surveyed, it must be remembered, is that of behavioral science, not "library" science. This process is known as the literature survey, that is the controlled analysis of previous work done in the field. The produce becomes a state of the art, or that chapter of the eventual research report which is a survey of previous work in the field. Notes are taken and analyses are made upon which eventually one can base assumptions, hypotheses and definitions that will result in measurement devices. Conceptualization and categorization grow out of the analyses so that scaling can be developed: nominal, ordinal, interval.

The literature survey also provides a justification for doing the research study, a validation of the logical argument which is research design. The validity of the assumptions rest upon the literature survey and analysis. The hypothesis should be explicated into perhaps a number of subhypotheses. Some discussion should occur as to the relation of the hypotheses to the assumptions and to the justification of the problem analysis. Normally this should constitute a chapter, for example in the eventual research report which is different from the literature is composed of research studies, surveys and experiments in the field, rather than the myrid how-to-do-it reports that seem to clog every professional field.

Validation is an extremely fundamental question. It comes from the professional concensus as to the general and specific principles which govern purposes, policies and procedures. Validation is a deductive, axiomative argument rather than an inductive one. Verification on the other hand is an inductive assembling of data which is employed to verify or negate the hypotheses. Of course the scientific method includes both processes, but the validation function is a product of professional knowledge that has been organized into a theory.

Theory however does not directly result from research. A theory must be developed before research can be undertaken. In the process of research it is substantiated, revised, or negated. Theory is generated through the creative insights of a single researcher or small group of colleagues working closely together. Theory is a creative process the product of which is submitted to the principles of the field and validated by a deductive, axiomatic argument. The theory is eventually put to the test in the scientific method.

A theoretical position is essential whether obtained from the infrastructure fields of information, library and media science, or from a communicative profession, or possibly supportive discipline. It is quite appropriate and even essential to borrow theory as well as the findings of research studies that have been done elsewhere. However this does not relieve the communications researcher from the responsibility for an analysis of the literature in information, library and media science however weak in behavioral studies these infrastructure fields may be.

After the problem has been delineated, the hypotheses stated and the justification developed, then the next step is a consideration of measurement. The purpose here is to explore the relationship between definitions and the scaling devices to be employed and which guide the data collection. This raises the question discussed in the previous chapter of the quantification of ideas or concepts to be investigated. Quantification is in a way the numeralization of ideas so that eventually statistical analyses can be employed to rule out the possibility of chance as a factor in the logical argument. Once chance has been ruled out, then a judgement has to be made as to the relationships which occur. If the communicator has gone through the process of validation then he can risk his integrity in making some pronouncement (observation or judgement) which may be taken as new knowledge.

The question of measurement cannot be overemphasized. Simple enumeration may be a process of counting, but it is a process of counting instances recognized as having certain properties in common. Inference is dependent upon recognition of resemblances. An inference from resemblances is an inference from analogy. There must not only be repetition of instances, but also resemblances and variety in the instances themselves. The four measurement scales (nominal, ordinal, interval, ratio) attribute to non-numerical data some of the properties of numbers. Unless there is a symbiotic relationship between content and number properties, statistical manipulation will remain just statistical manipulation.

Nominal scaling is the process of identifying concepts and of categorizing them in sets which lack fixed interval boundaries between each of the concepts. For example, the dewey classes represent a nominal scale. Each class has a boundary and includes as a set all the subclasses or subideas. However there is no necessary progression in such a nominal scale. There is no "before and after." It is simply a happenstance for example that the content of class 100 comes before class 200 in dewey. A nominal scale is sometimes represented by a pie-chart, rather than a bar-chart or histogram. In this regard, patrons might be better served by open, readers interest shelf arrangements than the pseudo-ordinal display of dewey. But even Dana's exhortation on the value of museum work (Gloom of the Museum. Elm Tree Press, 1917 ), long ago fell on the deaf ears of librarians. Only recently has it been "exhumed" by media specialists for its important contribution to communications design.

Scaling is a process essential to the statistical analysis. Statistical analysis is mainly a method which rules out chance. Once the factor of chance has been accounted for, and there still appears to be a relationship between the independent variable and behavioral change, it then becomes necessary to make a judgement, or statement of explanation based upon the theory. In effect, the researcher is saying that the sample is a relizble predictor of the population on whatever characteristics he has undertaken to study. Since the sample behaves in a certain way it is safe to extrapolate and impute that behavior to the population from which the sample was drawn. If however the initial theoretical position does not account for the change actually observed, then it is necessary to draw conclusions that will assist in the reformulation of that theory into new principles.

Observational Design:

Underlying all research design and method is the assumption that some order exists, otherwise there would be no reason to seek uniformities. Human behavior is considered to be cybernetic and part of the order of nature along with organisms, galaxies, or transportation systems. New findings are sought in order to add to the corpus of knowledge primarily for its own sake and secondarily for particular social purposes if there is a socially relevant profession to make it kinetic in the lives of people.

Scholarships or knowledge is the cumulative system of research findings which have been used to establish new facts, their interrelationships as well as the power of their predictive value. The research orientation is characterized by openmindedness on the one hand, ie. a willingness to accept anything, but on the other by an extreme skepticism or reluctance to accept any fact without demonstration. In any event, a number of criteria are used to evaluate the contributions of any research study:

> Methods, procedures and results are both public and communicable in a detailed report of what was done and how in order that another competent researcher could follow each step in the investigation.
>
> Definitions employed are precise enough to be laid out in clear procedures. Data collection is so objective that another researcher can replicate the findings.
>
> Purposes should emphasize explanation, understanding and prediction so that the contribution will be systematic, cumulative and thus related to some segment of knowledge.

Observational methods are closely related to the logical argument which the researcher has proposed for establishing new knowledge. His argument is based upon both inductive and deductive logic. The argument has been validated on deductive logic and verified by inductive logic. Development of the argument moves back and forth between deductive and inductive considerations so that an appropriate mapping occurs between the two. In one sense, research design is a compromise held together simply because the areas of agreement outweigh the areas of conflict.

Research methodology is neither mysterious nor intuitive. Each method is a well ordered convention in the scholarship of western civilization for the handling of data. Methods vary over time and by place of origin. Witness the prevalence of the scientific method in the west and in particular the use of opinion research in advanced democratic countries. Nor is methodology unchanging and "sacred." Methodology is under continuous development. The beginning researcher should be cautioned that before he creates new methodologies for the present he should understand those of the past.

There are in general two basic methods of observations and data collection: the survey and the experiment. The difference between them is not so much one of kind, as it is variant emphases of man's curiosity and his need for assurance. The scientific method includes both processes, that of discovery as well as that of justification. The discovery phase includes the processes of intuitive insight, survey analysis and descriptive statistics. The phase of justification builds upon discovery analyses and moves into such emphases as rigorous logical arguments, experimental design and inferential statistics.

The process of discovery includes both the present moment slice-of-time survey as well as that of probing into historical antecedents. The phase of discovery emphasizes philosophical inquiry and reflective thinking on levels of generalization such that the characteristics of numbers cannot as yet be attributed to its data with any finality. Discovery often proceeds from the gestalt and cannot as yet find the broad generalization mirrored in the grains of data. Relevance occurs in the area of theory development and in the testing of the axioms underlying various disciplines. Models are sometimes developed and employed for the purpose of analysis.

Historical research deals with past experience in a manner similar to that which the current survey does for the present. Its method is generic and applies reflective thinking to social problems which remain unsolved by discovering past trends of event, fact and attitude. Consequently the continuity survey has an historical point of view, as well as the survey method which checks status at different points in time. In historical knowledge, primary facts are not observed but inferred from the records in which they are described. Documents may be collected, classified and annotated as a result of the work of many people over a long period of time. Each document must be understood and evaluated as a record of fact. The document is submitted to external and internal criticism.

As the empirical sciences have grown out of commonsense knowledge, so history has grown out of tradition. Just as science is something more than organized commonsense, so history is something more than tradition. Tradition becomes conscious and critical of itself. The historial method studies developments in space and in time whereas the sociologist is interested in generalizing about the quantity and repetition of the present. Both insist upon reliable observation. The historian however does not make his own observations. Historical method is in part the attempt to test the truthfulness of reports made by others who are usually not trained observers as in the case of the sociologist. Therefore formal records are combined with the recollections of participants and eye witness accounts for greater accuracy.

Hypotheses are essential in well-ordered historical research in order to guide bibliographic activity and source consultation. Reading, guided by hypothesis, becomes definite and purposeful. Reading is in part a critical review and as such may be considered an aspect of content analysis. Source criticism is essential to historical method and deals with the value of documents. Initial criticism attempts to detect the spurious by forming judgments on the probable circumstances of document origin through internal analysis and a comparison with other sources. Textual criticism, a step higher, seeks to discover corruptions and changes which may have occured. Finally, internal or higher criticism aims at a final judgment on the meaning of the data gathered and at an interpretation of the statements which emerge from textual analysis.

Descriptive research deals with the present status of variables in any environment. A cross-section check of a situation aims to isolate problems and to obtain generalized knowledge for future use. Hypothesizing about solutions to the problems may lead to more rigorous research. If periodic reports of status and change are made, the method has moved into the genetic continuity study. If the identified variables can be controlled so that experimental factors can be isolated for measurement and change, then true experimentation may occur. Descriptive research may be defined as fact finding with adequate interpretation. Descriptive research exists in a variety of forms:

> Surveys help to analyze, interpret and report on the present status of a social institution, group or area. The purpose is to get groups of classified, generalized and interpreted data for guidance of practice.
>
> Continuity description is a survey at periodic intervals.
>
> Case study is a survey which aims to be complete about an individual, an agency or institution.
>
> Job and activity analyses are related to operations research.

> Library and documentary analysis are a survey and critical analysis of data in printed form (an aspect of historical research). They constitute a critical evaluation of a group of materials with interpretation.

Obviously there are certain questions that can be answered only by a survey. The question, "To what extent do American librarians today believe that communication services are necessary?" can be answered by specifying a population and then asking a selected sample. No experiment will answer the question and neither will a case study. Whenever the investigator is interested in assessing or estimating the present state of affairs with regard to some variable that changes over time for a large group of subjects, a sample survey is the only practical way to get the answer. If the variable did not change over time, we could probably learn the answer once and for all by experiment. If there was interest in only one or few instances of it, case studies could provide the answer.

A sample survey, as a research design, does not refer only to a public opinion pool. However a properly designed poll is one example of a sample survey, and probably the most familiar one. A sample survey is properly named when it contains the indicated two elements:

> A sample: The investigator first decides what group or population he is interested in (adults, voters, women of childbearing age, college students) and then selects a sample in the statistical sense. It may be "random," "representative," "quota," "weighted," or any of a number. The sample is so chosen as to enable the experimenter to draw conclusions regarding the entire population and not simply those members of the population who happen to turn up in the sample.
>
> A survey: The investigator then collects some measures on the appropriate characteristics of the population being studied (number of television sets or children in the household; how the members feel about the library or the church; what they know about library use).

In addition to simple measures of magnitude (how many people will borrow books?), surveys provide clues to relations between variables (and thus by extrapolation to cause and effect) by correlation of the various measures obtained. For example, a survey of the number of books borrowed per family can provide a series of tables showing how information center use varies by families of differing class, race, rural-urban residence, or religion. This example, incidentally, illustrates another advantage of the survey in the study of relationships. Many times the variables of interest are difficult or impossible to manipulate by experiment (years of schooling, race) so the only approach is to compare people who already differ on the characteristic in question and see how their behaviors differ.

Such correlations are difficult to identify causally, because the direction of the influence is uncertain. It is often reciprocal, which makes the influence more difficult to determine. For example, a correlation between reading a review for a given book and buying the book could go either way: review reading influenced purchase, purchase influences review reading. Even when the direction is clear, when one characteristic (eg. socioeconomic status) antedates and is not affected by another (eg. library borrowing), the nature of the causal relationship is quite complex, with several other factors usually involved (eg. income, social position, religion, place of residence, marriage).

To handle change over time in certain investigations, a variant of the sample survey has been developed that is called the panel. This requires repeated measures of the appropriate characteristics on the same people. Consequently the investigation can study how changes were brought about over time. This method is particularly useful in campaigns that bring a variety of stimuli to the subjects' attention. It is no accident that the method is used mainly in studies of audience research and market analysis. However, as the same people are queried repeatedly, they may change their behavior simply as a result of panel membership. As a control, panel responses are often checked against samples of "fresh" respondents.

The phenomenon of regression is an interesting characteristic of two rectilinearly related traits for any group of individuals. In any sample from a population a number of individuals may be exactly alike on one trait. However, these individuals will lie closer to the general average on the second trait than they will on the first one. For example, consider the relationship between height and weight. If from a large group of adults, a number of people 6 foot 3 inches tall were selected, this selected group would be much less extreme in weight. In weight, they would be much closer to the average of the population as a whole.

Another example may consider a test in library use and one in world history administered to a large college freshman group. From the total, a number of individuals who scored high on library use could be selected. The investigator would find that these individuals, who may be above average on the history test, would not with few exceptions be as far above average in history as in library use.

For these selected individuals, the mean score on the history test would be lower (when expressed in comparable terms e.g., z-scores) than their scores on the library use test. Similarly, most individuals making very low scores on the history test would make better scores on the library use test.

The case study complements the survey. The survey measures many people on few characteristics, usually at one point in time. The case study intensively examines many characteristics of one "unit" (person, work group, company, community, culture) usually over a long period of time. The goal of such investigation is to learn "all" about the area of interest for the one case involved. Case studies often serve as a major source of data on many questions in communications. The researcher may be interested in specific patrons, groups or communities and collects intensive data on these adaptive control systems. On the other hand, the researcher may be interested in the **behavior** of the professional communicator as he functions in each of these three contexts.

The detail and the depth of information over time that the case study provides makes this design particularly relevant for questions involving etiological development: How does a particular information use pattern emerge and change over time? What are the critical incidents that lead up to a library strike? How does the automation of a library affect the staff and the public? The chief limitation is that the results are based on a sample of one. Consequently the degree of their generality is not known. Would another individual, another library, another community, another culture respond in the same way?

Neither a "control group" nor any intervention by the investigator is provided for as a safeguard. Case studies rarely prove anything, although frequently they are rich in clues and insights for further investigation. In many areas the case study is the idea-getting investigation par excellence. However, the results of case studies can only be used for predictive purposes when the conclusions have been verified in some other way.

The process of justification includes both problem solving for decision making as well as the more rigorous experimental design. The purpose of problem solving is to control future operations of the factors investigated in order that the inevitable things which must be done or applied may be controlled more intelligently. It is closely related to decision-making and is based on any other type of research that is appropriate. It often results in invention rather than in controlled research. The problem solving method moves in a progression from the subjective and the unique to the objective and therefore replicable creation of new knowledge. At least four levels in the progression may be identified:

> The stylistic method employed in the humanities determines the aesthetic qualities of the object or situation. The result may be emotional and aesthetic satisfaction. The factors and identities of intrinsic worth are selected to conform to personally held standards rather than professional ones. It often naively interprets concomitance as causation in human experience. In addition to the methods of reflective thinking it includes a strong, pervasive and continuous feeling tone.
>
> The telic approach is that of invention. There is a specific purpose or end in view. Generalizations are sought and used in a practical way leading to invention. The work of testmakers in education and psychology may serve as an example.
>
> The genetic method traces the history of phenomena. Reports of origin and trends of development result. However the validity of generalizations is assumed rather than investigated. The operation of generalization is observed rather than investigated. Interpretation is frequently on the basis of how fully the generalizations satisfy points of view held by the investigator.
>
> The experimental approach offers the highest degree of prognostic certainty. Factors can be isolated, manipulated and measured. Analysis reduces data to simpler components. Generalizations appear and eventually laws may be announced which summarize the permanent value of generalizations.

Sometimes the genetic method moves into the continuity study and historical research, the case group and the case history. These methods help to strengthen and tighten up on the assumptions that conditions will continue unchanged into the future. Another type of genetic investigation is the correlation study which traces the relationship of measured variables over definite periods of time. Trends thus revealed are recorded in terms of the regression equation which records the trends of value in the realm of concomitance investigated. Graphic representation displays a clear, understandable picture of trends. Description discloses the discovery of trends and records knowledge of the experience. Relationships are also recorded and summarized in descriptive statistical measures especially partial and multiple correlation.

Considerations of criteria in evaluating problem solving investigations include the degree of certainty which can be placed in the prediction. Of course, prediction increases with the sophistication of research procedure. Of equal importance is the criterion of value and the intrinsic value of the data for the largest generalizations possible. Problem solving investigations require careful definition and limitation to a worthwhile concern, a skillfully administered measurement program, good classification of data for ease of interpretation, a readable final report that makes appropriate decision making possible, and finally a definite provision for further research by carefully listing the related problems.

Experimental method can occur when variables have been identified and can be controlled. Consequently experimental factors can be isolated for measurement and change. It provides for the repetition of an event under controlled conditions so that an understanding may be gained of what is taking place. Adequate control is the essential factor. One circumstance can be varied at a time while all the other circumstances can be rigorously controlled. The five canons of Mill are classical guides to the justification and verification of experimental truth. The five canons can be reduced to two methods: compare and contrast instances in which one variable occurs; compare instances in which the variable occurs with identical instances in which it does not occur. In addition to these methods, seven prescriptions may be taken into consideration:

> Problem should be isolated from the vague feeling of need preceeding reflective thinking.
>
> Experimental factor(s) should be carefully defined.
>
> Apply techniques appropriate to the experimental group whether one group, equivalent groups or a rotation.

> Control all factors or variables operating other than the experimental variable itself.
>
> Select or make and administer reliable tests in a carefully planned measurement program.
>
> Report all steps taken and methods used from beginning to end of the project.
>
> Be prepared to repeat the experiment in another situation or with other control and experimental groups. At the least, provision should be made for further research in the same or related fields.

Experimental design proper is any investigative design which includes: manipulation or control of some variable, as well as the systematic observation or measurement of results. The experiment is an active intervention in some phenomena of interest to see what effects can be produced. Experimental design, at least in the social sciences includes in its approach a broad view of human change and measurement. For example, an attitude is considered to have been measured if such dichotomous categories as "for" or "against," as well as "more" or "less" can be determined.

The prototype of experimentation is the classical experiment. The general question it answers is whether, and to what extent, one variable (experimental or independent variable) affects another variable (dependent variable). The logic is simple and "foolproof." Two groups are matched at the outset. One is given the experimental intervention (library public relations, propaganda, a new book that affects behavior, a training lesson taught in a new way, a special procedure introducing changes in work flow). The result of the intervention is subsequently measured (effect on attitudes, personality, amount learned, morale and productivity).

The importance of experimentation depends not so much on its precision, its objectivity, or its instruments as on the inherent efficiency of intervention and in disentangling the mutual, or cause-and-effect relationships. Whenever its use is feasible, purposeful intervention is a method that most readily identifies cause and effect. If communication science was able to experiment more widely on human subjects, it would be better equipped with important findings. However, in order to offset the limitations imposed on human experimentation, model design and simulation have been employed as alternatives to the rigorous experimental design. In any event, the steps in the classical experiment include the following.

> Define the population of subjects and draw a sample. Other than in sampling error, the sample is representative of the population.
>
> Divide the sample at random into two groups. By definition the two groups are similar, within limits of sampling error, on any measurement. There is no reason to expect one group to behave any differently in the future than the other. Flip a coin to decide which will be "experimental" and which "control."
>
> Define the dependent variable. How will it be measured or rated? Take a "before" measurement on each group.
>
> Define the experimental variable precisely and administer it to the experimental group only. The control group may get a placebo to control for the effects of autosuggestion, and even for the effect of participating in the experiment). Where possible the experiment is "double blind". Not only does the subject not know which pill for example he gets but, in order to control his own wishes in the matter, the experimenter may not know at the time either.
>
> Take "after" measurements on the dependent variable on each group. Any differences between the two groups after the experiment, beyond any difference that may have existed before, is the effect of the experimental variable.

When all the conditions of the classical experiment have been met, the final difference between control and experimental group must be due to the effect of the experimental or independent variable. Both groups reflect in "equal" degree the effects of any other variables not directly manipulated by the experimenter such as time, atmospheric conditions, or the effects of being selected to participate. Thus the control group protects the experimenter against many of the common fallacies that plague less rigorous studies. Without a control group there is a temptation to attribute any subsequent change in the observed subjects to the experimental variable, whereas the change may have occurred despite the experimenter's intervention.

Despite the logical model of the classical experiment, the design is sometimes modified for various reasons of cost and practical difficulties. In some cases, statistical approximations suffice. For example, if the experimental and control groups are truly divided at random, the before measurement may be omitted in the knowledge that the two groups will vary only within known limits of sampling error. The experiment then consists of the experimental condition being administered to one group and the subsequent after measurements on both. If this after measurement records a difference between the two groups that cannot be attributed to chance variation, it is taken to be the result of the experimental variable.

When a population sample has been divided again at random into two ore more groups, the groups will differ by no more than chance on any characteristic whatsoever, be it blue eyes, redheads, or those who have taken a book from the library. All individuals within groups will be roughly equivalent. Mathematical procedures can be used to determine the probability of a given difference having arisen due simply to such random division. Therefore, when a difference is greater than that which could reasonably be expected on the basis of random division, and the groups have in fact been randomly divided, the conclusion is that the difference is not due to their division but to something that happened to them afterward (i.e., the independent variable).

Many experiments add onto the basic experimental model. Some measure the effect of the experimental variable over time. Propaganda may be effective right after its administration: but how long does it last? Other researchers assess the effects of several experimental variables within a single investigation. The simplest form of this involves two or more experimental groups, each of which is measured against a single control group. An investigation of library use learning might vary the method of instruction, the sex of the librarian instructor, the room lighting, and patron motivation, all at one stage of observation. Then the conclusion could be drawn as to which of these factors have the most important influence on learning, and how they act in combination. Modern statistical designs for example analysis of variance (ANOVA), make it possible to evaluate the relative effects of a number of independent variables acting simultaneously and in combination.

One problem of experimental work done in the laboratory is that of "translation." Some laboratory experiments deal with phenomena that can be reproduced directly in the laboratory, such as depth perception, small-group problems, or the learning of certain skills. In such cases, the experimenter may bring the behavior of interest into the laboratory for a more careful scrutiny under experimental conditions. After the experiment, translation of the results of the specific operations performed must be made back into the original concepts and phenomena. Many phenomena, however, cannot be transplanted to a laboratory either in principle or for such practical considerations as time, money, cooperation of subjects. In addition, the phenomenon of interest may be a general one (motivation, love, hostility) that must be delineated in a single specific instance.

Evaluation of such experimental results demands critical consideration of the translation involved and its validity. Similar considerations are necessary when the findings of research in one discipline appear to have relevance to another field. This imperative is especially pressing in translating the research findings of the disciplines into the context of the professions. Professional people often reject research findings not because the studies are poorly designed or failed to produce conclusive results, but usually because the findings appear to be irrelevant to their preconceptions.

A so-called natural experiment is conducted when the major elements of an experiment occur or are produced in the natural habitat of the behavior under study. Such experimentation avoids supposed problems of the laboratory situation such as oversimplification and artificiality. In the natural experiment, subjects ordinarily do not know they are under investigation and as a result presumably do not modify their behavior as a result of being watched. On the other hand, natural experimentation is usually less precise. The pertinent events are less fully under the experimenter's control.

The planned natural experiment extends the method somewhat. The researcher manipulates the independent variable to meet his purposes and then makes systematic measurements of the result. For example, this occurs when variant forms of an instructional presentation are administered. Another example of a planned natural experiment often used in advertising research is the "split-run" technique. Metropoliten newspapers frequently offer advertisers the opportunity to run two versions of an advertisement in the same issue. Coupons or other coded devices are appended to one version which enables "returns" from the two versions to be compared.

The "spontaneous" natural experiment occurs when the researcher is fortunate enough to come on a situation that happened by itself yet has most or all of the elements of a successful experiment. In such cases, an approximation of experimental results may be obtained. For example, when television was originally being introduced in America, there was a period when technical considerations determined which towns and cities would have stations. As a result it was possible to find a number of cities without TV and to ascertain when the medium would be introduced. This provided an opportunity to study "what television does" on a before-and-after basis, as compared with matched towns without TV (the controls). In a similar way, a particular library agency can be studied before and after the introduction of automation or the introduction of a communication services program.

In summary, some readers of this section of the manual may consider the approach to research methodology cavalier, or at best cursory. But the methods presented are traditional and therefore basic to an understanding of the newer communications research methods. The profession of information, library and media science desperately needs to bring itself into the mainstream of behavioral science research. Consequently, the purpose of this manual is to clear the way for experimentation and innovation in research design.

## Observational Techniques:

Research design requires that information about human behavior be gathered and recorded in some systematic fashion. The major methods for collecting data used in the behavioral sciences include: observation, report, records. It is possible for the researcher to observe the behavior he is interested in as it occurs. He can do this either as a participant in the event or the community which is being studied, or as a nonparticipant. In either case, the fact that they are under observation may or may not be known to the subjects as is evident from the figure: (Direct Observation.)

The relative efficiency of various observational techniques depends on the character of the problem. A basic consideration is the extent to which the phenomenon under study is likely to be modified if the subjects know they are being studied. In principle there probably is no problem in the behavioral sciences that is not affected at least potentially by the subject's knowledge that he is participating in the study. When people are observed by others, especially in behavior that is personally or socially significant, they frequently behave quite differently than they otherwise would.

| | Participant | Nonparticipant |
|---|---|---|
| Known to subjects | Researcher enrolls and goes through librarian's training program to learn how recruits are taught. Identity is known. | Researcher sits in a corner and takes notes, taperecords or watches through a window. |
| Not known to subjects | Researcher gets job as librarian in order to study behavior of library staff. | Researcher in library observes and records behavior of staff members in performing assignments. |

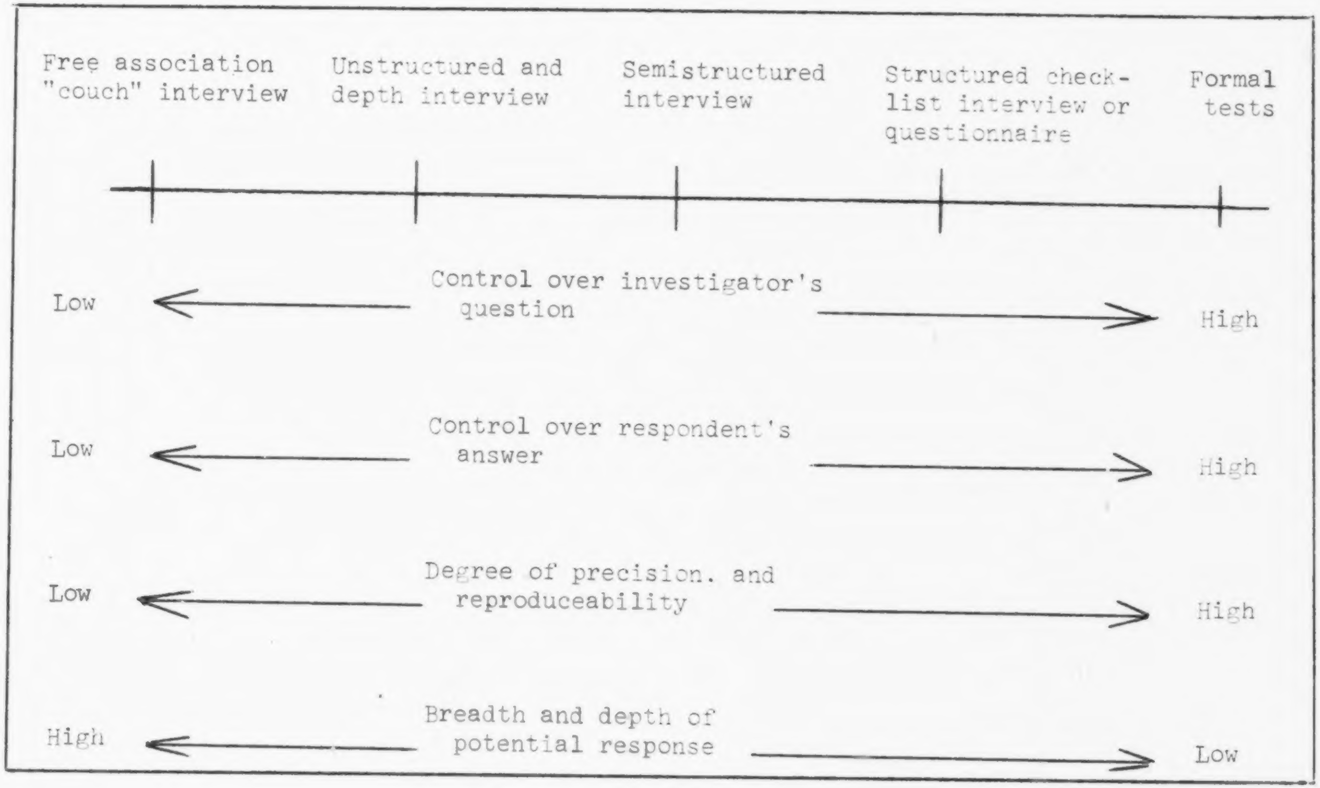
DIRECT OBSERVATION

People, at least in our culture, have a desire to please or impress, to be considered intelligent, acceptable or moral even by an impersonal observer who does not judge their actions as such and who is a stranger unlikely ever to be encountered again. This tendency on the part of a knowing subject may be irrelevant or insignificant in some studies. It can sometimes be controlled experimentally (as, for example, by a control group that also knows it is under study). But when the investigator wishes to describe behavior of a personally important or private nature, direct observation known to the subjects may be impossible. If any ethical problems of surreptitious observation are weighed, the researcher may wish to reveal his surreptitious observation and ask permission to use the data so collected.

The actual means of recording direct observations vary. They range from impressionistic field notes to precise and highly quantified ratings or precoded observation sheets. One technique quantifies all behavior that occurs in a dyad situation into a system of categories which are recorded, classified and analyzed. In another instance, the observer may be located behind a one-way mirror, where he can record on a moving belt the complete proceedings of a meeting, indicating who speaks to whom, in what order, and in which category each exchange falls. Direct observation also includes records made on films, tapes, and other mechanical devices. In many instances, such records can yield quite accurate and comprehensive data. The data can be studied at leisure by any number of observers (independently or in unison) and checked and rechecked as often as needed.

Verbal reports are often necessary to overcome the limitations imposed on direct observation. Some phenomena cannot be observed directly, such as behavior that is private (sexual behavior, arguments between husband and wife) or asocial (criminality and trickery) or protected by custom (certain religious matters). In addition, there is a vast amount of behavior that simply does not exist in directly observable form such as ideas, attitudes, feelings, beliefs. Finally, it is frequently not feasible to make direct observations in the number and scope that may be desired for a given study, even though it would be possible to do so.

Consequently, a variety of techniques already exist or can be created which can be used in obtaining and recording observational reports. These techniques are variants of the interview or the test and include: free association; the unstructured, semisturctured or structured interview; and formal tests (see figure: Observational Techniques). A great deal of data on human behavior comes directly from these kinds of reports. As more and more constraint is placed upon both the interviewer and the respondent, any unknown biases introduced by the method of interrogation decrease while precision and reliability increase. Respondents confine themselves more and more closely to a specific topic and their answers typically become more directly comparable one to another .


Free association "couch" interview
Unstructured and depth interview
Semistructured interview
Structured check-list interview or questionnaire
Formal tests
Low
Control over investigator's question
High
Low
Control over respondent's answer
High
Low
Degree of precision. and reproduceability
High
High
Breadth and depth of potential response
Low

OBSERVATIONAL TECHNIQUES

On the other hand, the potential range of response decreases as constraint increases. The more structured the interview, the less "broad" the information obtained. In any event, the interview can range in its progress from the initial exploratory stage, through the middle range where the investigator follows up hypotheses about his material, to the final step where he is concerned with particular points. Obviously in keeping pace with this progression, the techniques of interviewing would move from nonstructured through semistructured to totally structured. If structured techniques were applied in the initial stages, the risk is run of ruling out entire areas of response that did not occur to the investigator. Counterwise, if nonstructued techniques were applied in the final stages this would make it less likely that any hypothesis will be rigorously confirmed or refuted.

Free association represents the ultimate in freedom of response. In that context, the interviewer refrains from any direct questioning, the subject simply reports anything that comes into his mind. Even the constraints of speech are removed and the process approaches the content of dreams. The material in such a transcript is probably as broad and deep a human report as can be secured in the form of verbal data. The subject may cover an exhaustive series of topics. Each is treated at length and on repeated occasions. The whole transcript includes mention of behavior and feelings probably never before or again reported to anyone. But such interpersonal contexts are rarely recorded and even when transcripts are made they are restraicted to the analyst or a small group of his colleagues.

The unstructured or depth interview approximates maximum freedom for the respondent in discussing a particular topic. But it does confine him to an area of concern to the investigator. It frees the answer but limits the questions. The interviewer has only an outline of topics to be covered. The wording, sequence, and intensity of questioning are left ot the interviewer's discretion. Ordinarily, they are tailored to each interview so as to produce the most meaningful material with the particular respondent. Depth interviewing uses the "nondirective probe," questions designed to produce further elaboration without influencing the content of the response. Counselor librarians are probably quite familiar with the techniques.

The unstructured interview attempts to elicit all of the respondent's information and attutudes which are related to the topic. In some cases, the discussion is recorded verbatim, and the interviewer also notes revealing facial expressions and gestures. The recorded transcript permits an analysis of the emotional flavor and the meaning between lines which are communicated by the respondent's own language and subtle coloration of terms. Such data are not available when an answer must be given to specific questions in specific terms. It is these characteristics which are referred to as being "in depth," and not to the degree of private attitude revelations.

The semistructured interview gives greater control to the interviewer than does the unstructured. The structurization process specifies not only the major topic to be stressed but also the questions to be asked and usually their order. Questions are of the "open-ended" type ("What do you think of the library?"). The wording of the question may be specified but the wording of the response is left to the respondent. The semistructued interview is a compromise between the open and closed interview. It acquires some of the advantages of reliability, precision, and control associated with more sturctured techniques. It sacrifices some of the scope and depth of response obtainable by less structured interviewing methods.

In the structured interview, both the interviewer and the respondent are limited to the specific wording and sequence of questions. The respondent is also confined to a precoded set of answers, made available to him by multiple choice or some other device. "When do you think librarians will become communications leaders: this year, in one to five years, in five to ten years, never? Or don't you know?" Another familiar form occurs: "How do you feel about the statement that it's more important to advance our knowledge of people than it is to build bigger libraries: strongly agree, mildly agree, can't say, mildly disagree, or strongly disagree?" Perhaps the simple dichotomy is used: "Have you borrowed a book from the library in the past two months--yes or no?"

The difference between the test for research purposes and the interview is obscure, if indeed it exists at all. An ordinary paper-and-pencil, multiple-choice "attitude" test varies little from the characteristics of the structured interview. The term "test" usually refers to the subject's performance on a task as scored or evaluated by the investigator. The questions merely happen to be printed instead of spoken. Such research tests are in effect self-administered interviews and represent extreme control on both questioner and respondent. A variant of the interview and the test is the inventory. The inventory form asks a subject to answer fairly direct questions about himself (e.g., "How often do you have nightmares?").

Tests and inventories measure or evaluate the individual directly. whereas interviews (excluding psychiatric ones) secure information indirectly on the respondent as he reports on some event or topic. Consequently tests are conducted on "subjects," whereas interviews are held with "respondents" (giving information about themselves) or with "informants" giving information about others in the community, the society, the culture. Personality tests include both precoded paper and pencil tests as well as those in which the subject's free response is interpreted by inference. Free response may be interpreted by such common "projective" tests, as the Rorschack in-blot test or the Thematic Apperception Test which requires telling a story from a picture.

Historical records and artifacts may be needed when the material of interest is simply not available for an original investigation -- for instance, when it occurred a hundred or more years ago. More recently but still in the past, behavioral scientists have fortunately been building data banks of recorded observations. While originally collected perhaps for other purposes, the data is basic enough to support various other research interests. Traditional historical research had no control over its observational and data collecting methods, whereas the newer historical approach uses data banks that were deliberately planned and developed for future historical research.

Information recorded for other purposes but of great potential interest to students of human behavior comes from many sources: census statistics; life insurance records, medical histories; newspapers, magazines, novels and autobiographies; admission and release records of psychiatric and other types of institutions; records of prisons, churches, and schools; business transcripts of various kinds. Much of this kind of data is of a scope such that no single scientific investigation could afford to collect. It is hard to think of any recorded information regarding people and their behavior that could not be of potential value to behavioral science. The range of data compilation is limited only by the ingenuity and energy of the investigator and the quality of the record.

In order to retrieve data from the volume of published material, the method of content analysis has been developed as a way of quantifying the characteristics of interest that appear in communications. This method is used quite extensively, especially for descriptive purposes. It consists, in effect, of classifying the content of books, newspapers, speeches, television shows, documents, or any other communication material into a set of categories so that their relative frequency can be determined. Is there more library news in the papers now than there was in the 1930's? What shifts are made in adapting novels into films? What is the image of the librarian as it appeas in contemporary fiction?

Many observational techniques exist. After the communicator obtains an understanding of them and the principles of measurement design, he should be encouraged to experiment with additional innovations. In this as in other communicative endeavors, systems analysis methods are useful in design development. Some planning device may prove helpful such as that presented in the figure: Observational Systems. Based on purpose (ultimate objectives) each measurement technique should be identified and specifically related to the immediate objectives to be realized from its use in the proposed research design.

| Investigative Purpose | Instrument Name and Function | Type of Data Collected |
|---|---|---|
| Explore the range of skills and techniques in counseling, group and community development work. | Role simulations include situations comonly encountered by librarians whether in service to individuals, groups and communities. Each situation displays variant effectiveness in role behaviors. | In each of the three contexts. several stimulus situations are presented. Respondent checks one in sets of multiple categories (varies 5-7 point scale) of response for each context situation. |
| Determine leadership potential considered as an ability to analyze and work in undefined situations. | Intolerance of ambiguity as a standard sociological instrument which probes tolerance of ambiguous situations. | Respondent reads each statement and reacts to one in a multiple category (7-point scale) of response to 16 items. |
| Ascertain the understanding of creative and critical thinking as employed in problem analysis and solving. | Semantic differential is a standard psychological instrument which probes the semantic space of certain concepts. | Respondent reacts to each concept (not more than 15) in variant response catagories (usually 8-10) on a 7-point scale. |
| Probe the appreciation of and commitment to the functions and roles of a librarian. | Attitude instrument as a standard social science tool measures degree of commitment towards public services. | Respondent reads and reacts to each item (not more than 30) on a standard 5-point scale. |

OBSERVATIONAL SYSTEMS

## EXPERIMENTAL CONTROL AND EXTRAPOLATION

Since evaluation deals with values, and values are personal, evaluation involves subjective judgments. Two individuals evaluating the same thing may not agree as to its desirability. There can be no completely objective evaluation even though the data upon which it is based may be objective enough. Few people exist who may not have heard of the desirability of objectivity in the pursuit of scientific information. However, in actual practice most people are not entirely free of the tendency of permitting their subjective values to bias observations. Of course any emphasis on objectivity is in itself the expression of a value judgment and is being questioned today. While the rigidity of baconian objectivity may be suspect, it should be remembered that gestaltist movements have promoted greater sophistication in behavioral research.

The process of evaluation is one of classification. Occasionally it may be possible to classify in dichotomies: this thing is "good" while that thing is "not good." More frequently classification will involve degrees of "goodness." One thing may be judged to be "good" and a second to be "better." A third thing may be "poor" while a fourth is "worse." It may also not be possible to distinguish between the "goodness" of two things, in which case the conclusion is reached that they are "equally good" or "equally bad." While evaluation involves classification, it is possible to classfy without evaluating. Classification becomes evaluation only when some concept of value is included as a basis for the classes.

In some situations classes are defined in qualitative terms, that is in which (if any) of the classes each object (or individual) belongs. The object will either belong in the class, or it will not. Each class is discrete, and there is no question as to the degree in which each object belongs to the class. For example, a chair may be defined as something on which a person can sit which has legs and a support for the back. Objects may be classified according to whether or not they are chairs. The object either is a chair, or it is not. There can be no degree of "chairness." In addition, what used to be called a sofa can be included by definition as a chair. Also if a back-rest were attached to a saddle and the saddle placed on a horse, the horse and saddle would become a "chair." Obviously additional qualifications will have to be placed on the definition in order to make it classificatory meaningful.

In many cases involving classification the classes are arbitrary divisions of some characteristic that varies in degree. This is the same as saying that the variation is continuous rather than defined in terms of discrete class differences. With continuous variables individuals (or objects) may differ by extremely small amounts, and these extremely small differences are conceivable even when it is impossible except by instrumentation to observe their presence. Height is an example of a continuous variable.

The numerical data with which researchers work may be classified as either continuous or discrete. *Discrete* data are those which result from counting objects (or individuals) in discrete classes. With such data, fractions are impossible. The number of chairs in different rooms, registration figures for a library, the number of children in families who borrow films are examples of discrete data. Continuous data result from measuring characteristics which vary in amount. With such data, fractions are not only possible, but all possible fractional values conceivable exist. Height, weight, age, intelligence, and educational achievement are examples of continuous variables. Discrete data are the result of enumeration (counting). Continuous data are the result of measuring.

Measurement is a tool in evaluation. Measurement becomes evaluation whenever the characteristics measured have values associated with them. For example, in education most of the characteristics with which the researcher is concerned are of interest because people generally hold them to be of value. Therefore, educational measurement and educational evaluation are frequently used as synonymous terms. The importance of distinguishing between the terms comes from the fact that measurement is rarely an end in itself. Measurement is important largely as a means of improving the accuracy of the "value" judgments which support theories and eventually scientific laws.

The general purpose of evaluation is to determine to what extent objectives have actually been achieved. How successful has the communicator been in bringing about the changes which he sought in setting up and carrying through his communication program? What skills, attitudes, and behavior changes did he expect to accomplish after the program was over? Changes may have been expected to occur as a result of the presentation experience (i.e. research or independent variable as presumed cause) in the sense of doing something differently than the population or its sample did before, or having a general way of looking at things that were different than before, or knowing something that was not known before.

Basically, evaluation results from a measurement process. The way in which evaluation takes place includes some form of measurement or a series of measurements. These enable the measurer to tell just what effects his independent variables or experimental programs have had. Measurement techniques range from very crude and informal tools to relatively precise formal ones. At the simplest level of evaluation, one simply can observe an audience and see whether it seems to be paying attention or whether it's wandering off and going to sleep. At the most formal level, one can apply a complex experimental design in which before and after measures are taken. Such measures have been carefully standardized in advance and statistical techniques have been applied to the analyses of results to determine the extent to which changes are real.

Sampling:

Sampling from a population saves time and effort. In sampling only some of the people or items are measured, rather than all of the items involved, no matter what the unit being measured. Thus, instead of having to evaluate all the contents of every program on every person, the application of sampling ideas and principles allows the researcher to limit such measures only to a fraction of the people or material.

For example, the problem may be taken as to whether or not the bottles of glue coming off an assembly line in a library supply manufacturing plant are all right. One could open and test every single bottle but that would ruin the product. Every new bottle need not be tested and then the question must be answered: "How many should I take in order to be sure?" Whether the quality control analyst takes one out of 50 or one out of 200 or one out of 1,000, depends substantially on how many bad ones he is likely to find.

If practically none of them is bad, he will not need to take very many. But if bad ones appear fairly often, or if there's a wide variety in the end product, he must sample a fair number if he wants a proportion of all of them that are of a certain kind. It also depends, obviously, on how sure he wants to be -- what the consequences will be if a bad one gets out. If he wants to be very sure of this, he must sample at a high rate. If he is willing to tolerate certain amounts of spoilage, then he can sample at a much lower rate.

Consequently one of the first considerations that one must take in sampling is the necessity of getting a sample that stands for, or truly represents the universe from which the sample is being drawn. A universe can be of anything -- catalog cards, bibliographies, films, videotapes. A universe is simply 100% of anything being considered by the researcher and of course must be defined before a sample can be pulled from it. Failure to do so is generally a common error among people who start out to do some evaluation.

Another concern is drawing a sample lies in determining whether the sample represents the universe. Many ways exist in which to draw a sample so as to maximize its representativeness. A principal criterion is to ensure that each member of the universe has the same chance of falling into the sample as any other member of that universe. In other words, if a sample is being drawn of adults in the city of Pittsburgh, then any given man or worman 21 years or over residing withing the city limits of Pittsburgh should have exactly the same chance of falling into the sample as any other individual in the city of Pittsburgh above the age of 21: no more, no less.

Provided the chances are known, subject selection does not have to be exactly equal. There are times when this becomes important. On occasion, it may be important to investigate some characteristic in a sample of a relatively rare kind of person. For example, the researcher may wish to investigate the behavior of the older person faced with the surprise value of information. If the older person were allowed to represent themselves in their proper population proportion, very few of them would be obtained unless the sample from the general population was very large. Probably the researcher would decide to give older people twice as much of a chance or more of coming into the sample as anybody else.

This is an acceptable procedure, provided the researcher knows what the chances are of obtaining older people. When their chances are doubled of falling into the sample, they are simply cut down when the researcher wants to talk about the whole city of Pittsburgh. He gives each of the older people a half-vote so to speak. If they each have a half-vote, then they are not over-represented when it comes to talking about the whole population. When it comes to studying old people alone the researcher has enough of them so that he can study them as a subgroup without having quite as large a sampling error as might occur had a double sample not been taken. In fact a triple or ntuple sample can be drawn as long as the chances are clear and the researcher remembers to weigh them back when necessary.

Samples that do not guarantee equal selection of every member of the universe are known as biased samples. "Biased" is often considered to be prejudiced or in favor of something or other. But actually "biased" is a statistical term and means that some members of the universe have more chance of appearing in the sample than others. The method of drawing the sample is biased in favor of some people in it rather than others. It is not hard to draw a biased sample, so the researcher needs to avoid pitfalls in sample drawing that leads to undetected biases. One way is to use a random method of selection. From a list of people that constitutes the population or universe from which a sample is drawn, every n-th person can be taken off that list in a random way.

For example, the researcher may want to take every tenth name off a list. The question is whether to start with the first name or the ninth name or any other name. In order to give everybody exactly the same chance, the problem can be solved by taking ten slips of paper and writing a digit on each. The papers are folded and shaken up in a hat. Someone picks one of the pieces of paper out of the hat without looking and the sampler starts with whatever number is picked. Every single number had an equal chance of being picked. This method is the classic one and is actually the so-called "stratified random method," called that because the list is arranged in a certain order in advance, before a choice is made from it. The researcher wants to preserve that order, sometimes at least.

Suppose the researcher wants to get exactly the right representation of men and women. There are, say, a hundred people to choose from and a sample of one in ten is needed. That should yield five men and five women, if there are axactly fifty women and fifty men in the universe. But just by chance alone sometimes the sample might end up with as many as seven women and three men or vice versa. The researcher wants to avoid this situation and still use random methods. One way to guarantee getting exactly the right proportion is simply divide the universe into two subuniverses: fifty men and fifty women. A sample is taken at the same rate from each subuniverse: one out of ten on a random basis of the women, and the same from the list of men. Exactly five men and five women constitute a random sample of ten people.

A variant of this method can be employed to sample more than one thing at once. The researcher may want to get a proportional representation of men and women as well as young and old people. The group is divided again in half: the lower and upper half of the age grouping. Now the researcher has four subuniverses: old men and young men as well as old women and young women. The value of this method is evident. With the four sub-groups all exactly represented in their right proportion in the sample, there are just the right number of them in the sample as compared to their position in the total group or universe. At the same time the principle has been preserved of an equal chance for everyone in the total group or universe to appear in the sample.

The idea of stratifying has to do with arranging populations in order to reduce sampling error. On the other hand, the practice of clustering involves sampling in groups for administrative convenience, even though it tends to increase error. Instead of drawing individuals, it is possible to draw groups or clusters of individuals. This avoids an enormous amount of travel in case people drawn completely at random from the list are spread out over the country. In this method, individuals are arranged by geographic locale. A city can be divided into sections and every n-th household taken within every n-th block.

Suppose that from a city of 100,000 families, a sample of 1000 is to be taken. This means drawing a sample of one family out of every 100. Every family in this city should have exactly one chance in 100 of falling into the sample, and would if every 100th family were taken from the city directory. Again 100 numbers would be placed in a hat for the first draw. But this method would probably result in having to travel to every corner of the city. Suppose instead of that -- for purposes of convenience and cost-saving -- the city were divided into sections. As a matter of fact, stratification and clustering would be combined by organizing the city in socioeconomic districts. Usually cities are organized by the best neighborhoods, the not-quite-so elegant, and so on down to the poorest ones.

Within each of these neighborhoods the overall sampling rate can be applied to the neighborhoods arranged by blocks. Every tenth block is selected and every tenth household on a random basis within each of those selected blocks. Within the ten percent of the blocks selected every tenth household in every tenth block would appear. This would give the overall rate of 1/10 x 1/10 or 1/100 -- the sampling rate wanted. This also preserves our sampling requirement of giving everybody the same chance. Instead of going to every block, or practically every block, in the district the researcher can now go only to every tenth block. This process is repeated throughout every district in town.

While this is a legitimate procedure, it has probably run up the sampling error a bit. For instance, it may turn out that all of a certain type of people are on two blocks. If the researcher happened to strike one of those two blocks he would be over representing that particular group of people. But, on the other hand, should neither of the two blocks fall into the sample (which is even more likely for actually they have only a small chance), then they will be clearly underrepresented. No matter what is done, there is going to be trouble with the proper representation of that group and this is going to increase error. It does not however increase the bias or introduce bias. It is actually impossible to know whether there will be few or too many of that group. Therefore the direction of error is unpredictable. If the direction of error is not forced by the sampling procedure, then no bias is introduced. There is nothing inherent in the procedure to guarantee in advance that any particular group is going to be over or underrepresented, only that it will have equal chance.

Sampling error is simply computing an answer to the question: how far off might a measure be, based on a sample, as compared to what it would have been if the whole population or universe had been taken. The sampling error varies with the size of the sample and with how different the members of the population are from one another on a particular measure. Suppose someone were trying to find out how many fingers there are on the human hand. The first problem is to determine how many human beings there would have to be in the sample before generalizations could be made with some safety to all human beings. Maybe there are some six fingered people and perhaps some four fingered people. But they are probably rare. The chances are that if a sample of two people were selected at random from the world population each of them would probably have five-fingered hands. In this characteristic in which people are similar not many samples would need to be drawn in order to predict it.

But suppose something is taken in which people are different, eg. how much they weigh. A sample of two or three individuals taken at random from the world population would probably result in getting rather large differences. Some people might weigh under 100 pounds, some over 300, and anywhere in between. It is obvious to say that there is a lot of variation in the weight distribution of people. Given the same sample size as in the case of the number of fingers on the hand, there is going to be a much larger sampling error. Predicting everybody's weight is far more hazardous from just a small sample. As the variation becomes greater, the hazard increases. As the characteristic being measured varies from member to member in a population, the sampling error for a given sample increases. Conversely, as the variation is reduced, or gets smaller from person to person, from item to item in a sample, then the sampling error, given the same size sample, will be smaller.

It should be obvious that the larger the sample one draws, the smaller the error is going to be in predicting the total universe from the sample. Perhaps however it is not as clear that the proportion of the universe which the number in the sample represents is not particularly important, but rather the absolute number itself. Does it matter whether the population of the world is two billion or three billion? Suppose a random sample of 100,000 people were drawn at random from all the people in the world so that everyone in the world would have exactly an equal chance of appearing in the sample. Suppose the average of these 100,000 weights were taken, the result would be very close to the true average that is, the average of all the people in the world. As long as sampling procedures are valid, the fact that a sample of 100,000 was drawn would guarantee that the average based on that sample would be dependable and would predict closely the average of the total population.

It is impossible to specify categorically how large a sample has to be. The answer is always other questions: How accurate do you want to be? How much time, effort and money are you willing to spend to be that accurate? In any event, as the number of cases in the sample increases, the sampling error drops. The principle is similar to the minimax/maximax argument of decision making. For example, the average height of all men in the world may be 68 inches. But there is always the proviso of plus or minus so many inches based on the dispersion. It is the index of "so many" which varies with the size of sample.

The odds may be pretty heavy that the true average height is somewhere between 67 1/2 inches and 68 1/2 inches for everybody in the whole world. There may be no point in spending any more time, effort and money in finding out the average any more accurately. But if it is necessary to be right within the nearest quarter inch, then of course, the sampling size will have to be increased enormously. As a general rule it turns out that the sample size has to be multiplied by four in order to reduce the sampling error by one-half, other things being equal. Techniques for computing sampling error can be determined from texts in statistics. In any event, as the number of items in the sample is increased, sampling error is reduced. As variability is increased from person to person on the trait being measured, the greather the sampling error. Conversely, as the variability is reduced, the lower the sampling error will be.

Data Manipulation:

The majority of studies in the social sciences use sampling techniques in which observations and measurements are made on a limited number of subjects. Any single fact obtained from a sample (such as mean, median, percentile, standard deviation) is almost certain to differ by some amount with the corresponding fact for the population from which the sample came. Such obtained facts should always be considered as only approximations or estimates for the corresponding population facts. Because of this, it is necessary to determine how dependable or reliable the estimate has been and thus accordingly qualify any generalization based on it. On the basis of such reliability statements it is possible to rule out coincidence and accidental relations as well as determine the real significance, if any, of the attribute under consideration.

It is common to take a few (one or two perhaps) random samples from a population for research purposes. However if a large number (one hundred or more) random samples were taken from the same population, the means of each of these samples will distribute themselves in a normal curve. The normal curve of the various sample means is called the sampling distribution of the means of the samples. The form of the sampling distribution is the normal curve, even though the total population from which the samples were taken may not be a normal curve. In effect the normal curve of the sample means can be more normal than the population distribution.

It should be recognized at this point that the population mean is a fixed quantity for that population. The population mean does not fluctuate from time to time or from sample to sample. Nor is the population mean necessarily distributed either normally or in any other fashion. The population mean is a fixed quantity, while the means of the various samples drawn from the population do fluctuate and do form a normal distribution. Statistical inference is concerned with the central problem of evaluating the characteristics of a population or universe which are evident from a sample drawn from that universe. A population is a complete group of things, phenomena or people having identified characteristics or attributes in common.

A sample is simply that proportion of a universe which is available for study. A random sample drawn from a population is a sample so selected that each value in the universe has an equal chance of being included within the sample. Certain assumptions have to be made about the population and its sample:

That values in the population are normally distributed: however, actual data seldom follow a normal curve exactly; and by the nature of the problem of inference, we know little concerning the universe about which we seek to generalize. Experiment and mathematical investigation have indicated that we can proceed with good approximation even though this first assumption is not rigorously met. The most that it appears to require in practice is that there be no evidence that the universe is extremely far from the normal form.

A random sample is large which means thirty or more values: Paralleling large sample theory is small theory which on the same principles permits inferences to be drawn when samples contain less than 30 observations. However there are added complexities.

A random sample is however small relative to the total universe: Whenever the sample is large relative to the universe, inferences based upon it will be more accurate.

Perhaps an example will serve a useful purpose at this point. A sample (more than thirty) is drawn from a population of adults and a standard I.Q. test administered. Upon analysis of the test scores, the mean is computed at 98.5 and the standard deviation ($\sigma$) at 1.2. Upon these two facts it would be hazardous to say that, if the total population of adults were tested, they too would have a mean of 98.5 and a standard deviation ($\sigma$ or s.d.) of 1.2. The best that one can say is that the facts of the sample probably lie somewhere near the facts of the population. In any event, some method is needed of determining how "good" the estimate is, i.e. within what distance of the population mean does the obtained sample mean lie? How confident can one be that the obtained sample mean is within any given distance of the population mean?

Since the population mean is not known, it is impossible to say where in the population distribution, the obtained sample mean lies. However, the obtained sample mean was drawn at random from the population distribution. According to a normal distribution, one can assume at the 1% level that the obtained sample mean is within 2.58 $\sigma$ (standard deviations) of the population mean. This confidence at the 1% level states that the obtained sample mean does not differ from the population mean by more than 2.58 x 1.2 = 3.10. Of course, this sampling error may be in either direction. The population mean may be 3.10 units higher or 3.10 units lower than the obtained sample mean. The location of the population mean may therefore range from 95.40 (98.5 - 3.10 = 95.40) to 101.60 (98.5 + 3.10 = 101.60). At the 1% level, one can be confident that the population mean lies within this range from 95.40 to 101.60.

Other confidence levels can be established and the corresponding confident interval can be laid out around the obtained sample mean.

The confidence interval may be increased to 3 $\sigma$ (standard deviations) and confidence would be established at the .3% level. On the other hand, the confidence interval may be decreased to 1 $\sigma$ (standard deviation). In this case confidence would be established at approximately the 32% level. Usual procedure establishes confidence at the .05% level (3.5 standard deviations), or the .01% level (4 standard deviations). Once the confidence level has been established, the researcher can then go on to extrapolate from his hypothesis about the population characteristics.

In any event, an infinite number of sample means are distributed in a normal curve around the mean of the population or universe. Of all possible sample means, 95% will fall within two standard deviations of the mean of the mean of the population or universe. Correspondingly, the probability is 95% that our one sample mean will lie within the interval $\bar{X}_p \pm 2\sigma_M$. The chances are 95 out of a hundred that our actual sample mean lies within two standard errors of the population mean. With the corresponding probability of any statement being correct, it is possible to state that the population mean lies within any given interval around the sample mean. This interval is known as the confidence interval. The probability is 95% that the sample mean falls somewhere within a range plus or minus $2\sigma_M$ around the population mean.

Suppose the sample mean is taken at the upper limit of that range. If that were the case then $\bar{X}_s - 2\sigma_M = \bar{X}_p$. Now suppose the sample mean is taken at the lower limit of that range. If that were the case then $\bar{X}_s + 2\sigma_M = \bar{X}_p$. Both of these statements are consistent with the statement that the population mean $\pm 2\sigma_M$ has a 95% probability of including a given sample mean. But since they are the extreme statements which can be made, consistent with this 95% probability, they set limits within which the population mean should lie in 95 out of 100 cases.

In a large number of cases in which a random sample is taken from a statistical population, the interval $\bar{X}_s \pm 2\sigma_M$ will include the mean of the population 95 out of 100 times. This means that in the long run we should expect to be right 95% of the time if we state that the population mean lies within two standard errors of the sample mean. There is a 95% probability that the statement will be correct when, in any one case, it is stated that the true mean of the population falls within this interval. There is a 95% degree of confidence that the population mean lies within $\bar{X}_s \pm 2\sigma_M$ as determined from the random sample in any one case.

The 95% confidence interval is only one of such intervals which could be chosen. Another might be the interval $\bar{X}_s \pm \sigma_M$. This would be an interval half as large as the 95% confidence interval, and would have a degree of confidence attached to it of but 68%. This is because there is a probability of only about 68% that the sample mean lies within one standard error of the population mean. This is a probability of but 68% that the following statement is correct: the population mean lies within $\bar{X}_s \pm \sigma_M$.

Still another confidence interval would be $\bar{X}_s \pm 3\sigma_M$. Under any normal curve the area enclosed by $\bar{X} \pm 3\sigma$ is 99.7% of the total area. This means that 99.7% of all possible sample means, for the size of sample drawn, will be within $3\sigma_M$ of the population mean. The probability is 99.7% that one sample mean will lie within $3\sigma_M$ of the population mean. There is a very high degree of confidence that the population mean lies in the interval $\bar{X}_s \pm 3\sigma_M$. This degree of confidence is represented by 99.7%. Such a statement could be wrong only 3 times out of 1000 if the confidence interval was determined in this way.

| INTERVAL | AREA UNDER NORMAL CURVE | DEGREE OF CONFIDENCE |
|---|---|---|
| $\bar{X}_s \pm \sigma_M$ | 68 % | 32 % |
| $\bar{X}_s \pm 2\sigma_M$ | 95 % | 5 % |
| $\bar{X}_s \pm 2.5\sigma_M$ | 99 % | 1 % |
| $\bar{X}_s \pm 3\sigma_M$ | 99.7 % | .3 % |
| $\bar{X}_s \pm 3.5\sigma_M$ | 99.95% | .05% |
| $\bar{X}_s \pm 4\sigma_M$ | 99.99% | .01% |

Various intervals based upon other multiples of $\sigma_M$ can be determined as illustrated in the above table. The degree of confidence in each case is given by the percentage of sample means lying within the chosen number of standard errors around the population mean. Something will be noticed here that is simply in accord with common sense. The smaller the interval (ie. the standard deviation) the less confident the researcher can be that the true but unknown mean of the population lies within it. Conversely, the wider the interval the more confident he can be that it includes the population mean. Anyone, on seeing the mean of a sample, could specify an interval around it and say: "I *think* the population mean falls somewhere in here." Also he could add: "If I took a wider interval I'd feel surer of it including the population mean. If I took a narrower interval I'd feel less sure."

The statistical argument may be based on common sense, but the use of such analytical methods permits the researcher to conclude objectively how certain he is that the population mean lies with a stated interval. In addition, he can say precisely how much more certain he would be if the interval were widened by a given amount. The ability to give quantitative expression to the degree of confidence with which such statements can be accepted is an outstanding contribution of statistical inference.

Despite these necessary concerns for the level of confidence and the confidence interval, most researchers go beyond them and hypothesize about the population mean. Given a certain hypothesis concerning a characteristic of the population, what evidence can be deduced from a sample concerning the truth or falsity of the hypothesis? In other words, given an hypothesis concerning the numerical value of a population mean, what evidence concerning the truth or the falsity of this hypothesis can be deduced from a random sample drawn from the population? How low must the probability be reduced of getting the observed difference by chance before the hypothesis is rejected as being untenable? What is the critical probability above which we should retain the hypothesis and below which we should reject it?

Only four possible decisions can be made when an hypothesis is tested. Two errors can be made and two correct decisions. The errors are: reject a true hypothesis (type I error); accept a false hypothesis (type II error). The correct decisions are the converse of the errors: accept a true hypothesis; reject a false hypothesis. In any case, the decision made can never be absolutely correct or false. There exists only a probability that it may be so.

The researcher may adopt the 5% level as the critical probability. In this instance, he accepts the hypothesis when the probability of getting the observed difference exceeds 5%, and rejecting the hypothesis when the probability is less than 5%. This amounts to accepting the hypothesis when the discrepancy of the sample mean is less than two standard deviations, and rejecting the hypothesis when the discrepancy is more than two standard deviations. Using this critical probability the researcher would make an error of the first kind 5% of the time because even when the hypothesis is true, 5% of all possible sample means lie farther away than two standard deviations. Whenever by chance the researcher gets one of these (5 out of 100 times) he would make the error of rejecting a true hypothesis.

On the other hand, the researcher may adopt .3% as the critical probability, which corresponds to a discrepancy between the hypothesis and the sample mean equal to 3 standard deviations. When such an hypothesis is in fact true, only .3% of all possible sample means would lie farther away than 3 standard deviations. The researcher would make an error of the first kind only when by chance he happened to draw one of these. This is the same as saying he would make an error of the first kind only .3% of the time. The proportion of cases, in which an error of the first kind that of rejecting a true hypothesis, can be made as small as desired simply by reducing the value for the critical probability. In fact the percentage of cases in which an error of the first kind can be made is equal to the critical probability adopted.

However an error of the first kind is not the only kind of error the researcher has to worry about. There is also the error of accepting a false hypotheses. The lower the value set for the critical probability, the fewer the hypothesis which will be rejected. But in so doing the chances are increased of accepting a false hypothesis. It is impossible to state the percentage of the time when an error of the second kind will be made on the basis of any given critical probability adopted.

The chance of accepting a false hypothesis depends not only upon the critical probability, but more particularly upon how far away the false hypothesis lies from the true mean of the population. In the long run, in instances where hypothesis are false, some will be farther from the truth than others. It is impossible to predict the percentage of instances in which a false hypothesis will be accepted. The chances of accepting false hypotheses are necessarily increased as fewer hypotheses are rejected due to the use of a lower critical probability.

Data manipulation depends upon the quantification of the observations made of immediate research objectives. If observations cannot or will not be objectified (because of "humanistic" rationalizations) hypotheses cannot be controlled and tested in any rigorous fashion. Both descriptive and inferential statistics depend upon quantification. The index number accepted for the level of confidence in any decision made about the hypothesis is a computed statistic. The computational techniques available depend upon such descriptive statistics as the mean, standard deviation and sometimes correlational analysis.

In quantification, the researcher can be as primitive or as sophisticated as he chooses to be, or as he is permitted by his professional area of concern. For the communicator however, the level of sophistication in the social sciences is such that he avoids quantification at the peril of being irrelevant both in applying research findings and in creating new ones. It may be that the sociocultural climate in which his patrons exist eschews quantification. But for the communicator to be trapped in such an environment is to fail to understand people and how to help them other than as an occupation or with the "dogoodism" of a social missionary.

The basic objective techniques of statistical inference are the tests of chi-square, means and analysis of variance (ANOVA). The chi-square test of significance is used on the differences among the proportions of subjects, objects and events that fall into nominal and ordinal categories. This test is sometimes called a distribution free statistic because the observed data does not need to follow a normal curve. The means test on the other hand is used on the differences between two or more means which can be computed from normal distributions.

Inference:

Inferential statistics is the process of generalizing from a sample to its population. The population is the totality of all possible subjects (ie. persons, objects, events) which constitute the whole of anything which is being studied. A random sample is a portion of that population which is selected for study and analysis. Inferential statistics is the procedure by which the researcher infers (estimates) characteristics of the population (parameters) from characteristics of the sample (statistics) which he has studied and analyzed.

The word "estimate" is used to describe the isomorphy (the same as) between the characteristics of the sample and the characteristics of the population. The isomorphy is never perfect, never an absolute fact. The isomorphy is always a probability or an estimate. The researcher can be either "right" or "wrong" in his estimate or judgement about the isomorphy. In the final analysis, it is the researcher who must make a judgement based upon logical strategy. The strategy that is often employed is analogous to the mimimax/maximax strategy of decision making in operations research.

In research design, this strategy is called the argument from the null hypothesis. In this strategy, the researcher states for argumentative purposes that there is no relationship for the variable(s) studied between the two prototype samples (control and experimental) selected at random from the population of subjects. He even goes so far as to state for the argument that any observed relationship is only the result of chance. This is the same as saying that the mean of the experimental sample will not range beyond the fourth standard deviation of the mean of the control sample on the same measurement scale (ie. a distribution that is at least an interval scale if not a ratio scale).

The observations made on any variable of either of the prototype samples (control or experimental) of the population of subjects (persons, objects, events) will yield data (frequencies) that can be either scaled or proportioned. Measured frequencies may have nominal, ordinal, interval or ratio distributions. But in order that a distribution may scale it must have at least interval frequencies which assumes that it also has nominal and ordinal value. This means that the data of a variable is distributed on a scale of equally appearing intervals in ordinal order.

Nominal and ordinal distributions do not of themselves scale, ie. no precisely regular intervals can at the moment be indicated in the distribution. This is not to say that such distributions will never scale. As the result of more survey analysis in the discovery phases of research development, intervals in the distributions may be discovered. On the other hand, an interval distribution may be a logical absurdity as is the case of the dichotomous data obtained from the tossing of a coin. In any event, nominal and ordinal distributions remain as proportional differences and not scaled differences.

The concept of the measured distribution and especially the type of distribution is not academic. In either instance, the probability of variation due to chance is estimated in order to make a judgement based upon the type I/type II error strategy. However, the tests used to establish the level of confidence are different. Proportional differences are analyzed by the chi-square test, and scaled differences by the means test or one of its variants. The difference here is that between the discrete data and continuous data in the previous section of this chapter.

The minimax/maximax aspect of the null hypothesis argumentative strategy is called the dilemma of type I/type II error. The dilemma results from the researcher's indecision as to which standard deviation or fraction thereof, for example he will accept as a significant difference between the mean of the control sample and the mean of the experimental sample. This dilemma is not academic, but arises from the confidence or lack of it that the researcher has in the normal distribution (normal curve).

The normal distribution is an absolutely fundamental principle of all scientific research and has been derived from the mathematical law of large number distributions. There are of course other distributions which do not display data in a normal frequency curve. Logics other than inductive-deductive have been developed based on mathematical models to study and analyze non-normal distributions. In any event, the normal distribution is a fundamental prerequisite of inferential statistics.

The argumentative strategy of type I/type II error is at the core of research design based on inductive-deductive logic. The difference between the mean of the control sample and the mean of the experimental sample may have occured by chance even though the experimental sample mean may have moved beyond, four or possibly five standard deviations from the control sample mean. This statement that chance alone is responsible for so large a variation is called the null hypothesis. To continue to maintain that chance alone is responsible for so large a deviation is called the type II error. Conversely, to maintain that chance is not responsible for a deviation which is less than three and one half standard deviations in the difference between the mean of the control sample and the mean of the experimental sample is called a type I error.

In any event whatever statement is eventually made by the researcher, it has only a probability of being correct. The higher the probability becomes the more confidence the researcher can muster in the correctness of his statement. At some point in the probability, known as the level of confidence, the researcher can be fairly sure that his statement is correct but he can never be absolutely certain. It should be remembered that the formal argument of statistical inference as presented in research reports proceeds in a negative fashion. For example, a null hypothesis is set up. The researcher makes a type II error if he retains a false null hypothesis. Conversely, he makes a type I error if he rejects a false null hypothesis.

In any actual problem there are two dangers: that of rejecting a true hypothesis and that of accepting a false hypothesis. The danger of committing an error of the first kind can be reduced by reducing the value chosen for the critical probability. But this can be done only at the expense of increasing the danger of committing an error of the second kind. The final choice of the critical probability represents a compromise between these two dangers. It must be arrived at by balancing the consequences of making an error of the first kind against the consequences of making an error of the second kind. The more serious an error of the first kind, relative to an error of the second kind, the lower should the critical probability be placed and conversely.

When a man goes on trail for murder the supposition of law is that he is innocent. If the jury rejects a true hypothesis an innocent man is hanged; if it accepts a false hypothesis a murderer goes free.

Statistics cannot make a decision for anyone. They can only provide an objective and rational basis on which to make the decision. Statistical inference indicates how to reduce the chances of rejecting a true hypothesis (by lowering the critical probability) but in the process they warn that the chances of accepting a false hypothesis are thereby increased. Where the critical probability is set depends ultimately upon a judgment as to the relative degrees of risk the researcher is willing to assume of making these two kinds of error in an actual case. Two critical probabilities which are frequently employed by statisticians in practice are .05% and .01%. These index numbers correspond to observed discrepancies between hypothesis and a sample mean equal to $3.5\sigma_M$ and $4\sigma_M$ respectively.

There is one way of reducing the chances of accepting a false hypothesis without at the same time increasing the chances of rejecting a true hypothesis. This can be done by taking a larger sample. As the size of a sample is increased, $X_s$ will tend to lie closer to the true value for $\bar{X}_p$, as is indicated by a reduced value for $\sigma_M$. Within a given critical probability, errors of the first kind may still be made with the same frequency regardless of sample size. However, as $X_s$ moves closer to $\bar{X}_p$ it will in fewer instances appear consistent with a value other than $\bar{X}_p$, i.e., with a false hypothesis. Consequently by taking a larger sample, the chances of rejecting a true hypothesis remain constant as given by the critical probability.

When the probability is calculated of getting a discrepancy as large as or larger than that observed, it is usually done by adding together the two "tails" of the sampling distribution beyond the number of standard deviations corresponding to the hypothesis about the population mean minus $\bar{X}_s$. This practice is referred to as testing in both directions and is followed because:

> Since the hypothesis is formed before the sample is drawn, it is impossible to know in advance whether the observed discrepancy between the hypothesis about the population mean and $\bar{X}_s$ will have a positive or a negative sign;
>
> An observed discrepancy of given size would be equally harmful to the hypothesis whether it had a positive or negative sign;
>
> An hypothesis must not incorporate any of the information found in the very same sample which is used to test it.

It should be stressed again that statistical tests of significance are based on the fact that observed data possess the properties of numbers. The properties of numbers may be attributed to observations when the data have nominal, ordinal, interval or ratio distributions. While all of these distributions can be counted, only the last three can be ranked, and only the last two can be scaled in equal appearing intervals.

Descriptive statistics describe the properties of numbers which may be attributed to data once it is quantified in nominal, ordinal, interval or ratio distributions. Inferential statistics infers numerical characteristics of a population from the numerical characteristics of samples (whether control or experimental) which have been drawn from that population. Tests of significance estimate the probability of chance fluctuations and differ basically only on their application to data distributions.

The level of significance is the magnitude of difference between the mean of the control sample and the mean of the experimental sample. In communications research, the most commonly used levels of significance include the .05 level (ie. 3.5 standard deviations), and the .01 level (ie. 4.0 standard deviations). While the dictionary meaning of significant may be "important" or "meaningful" in inferential statistics the word means "less likely to be a function of chance than the independent variable." However, it must be remembered that research results can be statistically significant without being validly meaningful or important.

The inherent validity of significance should actually be as much, or more of a concern to the researcher as is statistical significance which is based on the properties of numbers. The logical canons of Mill constitute the validity of quantification. But the configurations of human knowledge were never built by researchers who had only a "nose for the facts." There are some considerations in answering the question of validity other than the canons of rigorous quantification.

The criterion of reasonable intuition specifies that if the facts (meaning numbers and the statements about numbers upon which statistical operations have been performed) conflict with what is intuitively reasonable to a responsible researcher, then the "facts" may be suspect. This criterion does not deserve the charge of subjectivity often leveled against it, but stems from the fact that although specified confidence limits indicated by mathematical probability do influence confidence the final decision is always the researcher's.

For example, the statement that ambiguity is a function of the audience is a remarkable insight into the communicative process. Such a viewpoint transfers to the audience a characteristic of information usually assumed to be predominantly a function of presentation devices, and thus assumed to be controlled or controllable at the source rather than at the receiver end of the communicative transaction. Perhaps if information specialists were aware of this intuition they might pursue the question of relevance elsewhere than in objective documents.

In order to avoid the danger of solipsism the researcher has to be an imaginative observer in his own and related fields of study.

Not since the time of Bacon have many scientists obtained concepts by the simple process of abstracting from empirical data. Concepts are rather intellectual creations which may be suggested by the data, but essentially are the products of constructive imagination. Consequently confidence in research findings may be related to the demonstrated competence of the researcher more as a constructively imaginative observer than one who is determined by his hypothesis.

It is a relatively simple matter to be blinded by one's hypotheses. For example, the communications engineer and the information scientist take a limited view of redundancy, and try to minimize it in order to increase transmission efficiency. For similar reasons, the English teacher takes a negative view of redundancy, identified as tautology, and expends considerable effort to teach students to avoid it. Actually, redundancy appears to be a requirement for human certainty faced with the surprise value of information. The researcher may be better advised to pursue redundancy as a function of behavioral outcomes than of message design.

Constructively imaginative insights applied to crude data, that may in themselves fail to satisfy the rigorous test of logical canons are at least a source for the advancement of knowledge. Isolated data, random points on a grid, and data compilations are of little use unless and until they are related to a systematic formulation which, by consensus, may be accepted as a theory. Confidence in the generality of research findings is enhanced when research findings are or can be related and integrated into accepted formulations of higher order abstraction, and weakened when this is not done or it is impossible to relate them to systematic formulations.

Replication of investigations of a problem may add to the confidence in the results, provided that the results of replications are consistent and extensive. In replication, neither the stimulus (or input) ensemble nor the techniques of investigation are sufficient, except possibly to satisfy some human need for compiled data. Nor, for that matter, is complete identity in replication likely to be achievable with human subjects. Replication involving a relative rather than an absolute constancy of the experiment is essential to increase the generality of the findings beyond the unique conditions of any one experiment. This criterion of replication leaves open the question of how many replications there need be and how varied the conditions.

The answers to these questions are functions, in part, of the individual researcher's tolerance of and criteria for risk. In any event, questions such as these and others can be answered in part by the clarity and logical argument of the research report. These objectives are explicated when the report follows the development and execution of a research design.

## SEARCH AND DISCOVERY

The researcher in information, library and media science may sometimes feel that he is familiar enough with subject analysis and indexing. He tends to consider any separate treatment of how to use information sources as superfluous. But in a surprising number of instances specialists in the field of information, library and media science browse at the shelf like the least sophisticated patron rather than in the subject file. Such a neglect to take advantage of the power of subject files should not be a characteristic of any researcher.

It may also appear to be obvious to remind communicators in information, library and media science of the necessity for and methods of doing a professional literature search. But since many communicators seldom do neither extensive nor involved searches and almost never in their own subject field, it may be useful to review the general model of searching for information:

. Analyze the subject: specific or general; part of what subject; related to what other subjects?

. Consult card catalog and note the books or classifications which might be useful (shelf-list may be used as substitute for card catalog)

. Check shelves for possible sources--examine books index; table of contents, illustrations, etc.; appendices; bibliographies.

. Systematic searching:
  .newest to oldest
  .use cumulated indexes
  .possible variations in spelling and filing
  .give complete bibliographical data in listing or citing references.

In reviewing any body of literature, there are in general three essential search strategies: work back from the new to the old; work down from the general to the specific; seek cognitive maps through coordination of terms. Of course, variations and combinations can be made of these into any number of specific retrieval flow-charts. In any event, the first search strategy is the simplest and most direct. It is employed almost exclusively with indexing and even abstracting services which do not have classificatory display.

Indexed information space exits in the linear temporal display which is typical of journal literature. In most instances the indexing thesaruus is not under authority control. Consequently, it is possible to study primarily the development of one or more variables and secondarily, in a more limited way, the relationships among variables.

The second search strategy requires that the indexing and abstracting thesaurus be under authority control. In order to search from the general to the specific, each term must be precisely defined and its relationship to other terms neatly identified. Any field or topic an be layed out in a logical and deductive taxonomy. The relationships among too or more variables are displayed based upon knowledge as created and organized in document pools. In addition distinctions can be made between the independent and dependent variables which have supported the creation of existing knowledge. Consequently, insight into intuitively new relationships may come most frequently as a result of reflection upon the product of the search retrieval rather than as a result of involvement in the process.

The third search strategy is more sophisticated than the other two and requires more instrumentation by the infrastructure technicians of information, library and media science. The additional instrumentation includes some indication of author viewpoint which makes it possible for the retriever to engage in an exciting and stimulating involvement with knowledge made dynamic. Traditional instrumentation has provided a fixed location class mark and this is reflected in document pools with classified shelf location. However where classification has been emancipated and documents are accessioned it is difficult if not impossible to engage in discovery retrieval unless an abstract is included on the unit card record.

Discovery retrieval is based upon the fact that it is possible to assign any author statement (document) to almost any subject in knowledge based upon the author's point of view. For example, an author treatment of "water" or practically any other topic for that matter can be assigned to any one of the ten dewey classes. The author's viewpoint is the determining criterion and an appropriate class mark may be assigned to that descriptor in the subject abstract. As the literature searcher compares and contrasts varying class assignments for an n-set of documents with a regressive feature appearing in the display of abstract descriptors (a reverse regression may as frequently occur), many questions arise which may develop into intuitive insights valuable eventually for research development and design.

The formulation of any research problem depends upon the identification and selection of the variables both dependent and independent. For example, experimental research consists mainly of determining the effect of independent variables upon dependent variables. The most common dependent variables considered in communications research may be included in these three categories: patron related; context (communicator) related; agency related. As might be expected, the agency related variables have so far received the most research attention in the infrastructure specializations of information, library and media science.

Patron Related Dependent Variables:

1. Patron appears before the communication librarian indicating some interest. Librarian conceives this contact to be one of a series and inquires whether the results of patron's previous contact were satisfactory.

2. Patron is given an initial counseling interview in which information about his interests, life goals and ability to use information sources may be obtained.

3. In those instances where the patron is too inchoate to discuss deep-felt concerns, audiovisual therapy or browsing in thematic displays may need to be used before proceeding to the second step.

4. At this point, in order to understand the problem-solving model of communication, the patron may need some instruction in the use of library resources as a method of thinking.

5. Employing the general problem solving model, the librarian together with the patron proceeds to enlarge the encounter to include outer-form materials which are more contextually oriented than subject classified.

6. When the subject and point of view have been established to the patron's initial satisfaction, a retrieval strategy (based on a boolean or other logic) is developed to search the descriptor file.

7. Document drops are examined in a "technical reading" for their information surprise value. If the information is not satisfactory to the patron, the search terms are regenerated and a new search strategy developed.

8. When a match has been achieved between the patron's goals and the kinds of information available, the librarian suggests related concepts and contexts in the exit interview for possible follow-up activity.

9. Patron enters upon a program of continuing development either in cognitive content mapping or in an n-dimensional matrix of interlocking situations and relationships.

10. Employing the method of case-load, the librarian is ready on a continuing basis to assist the patron to enter the communications system at any point and be expedited to any other point he may wish to travel.

11. Since many sources are available they have to be carefully matched to the goal to be achieved and to the task conditions for which they are to be employed. For example, pictures and motion may be substituted for concepts. Dialog and group discussion may enhance the intrapersonal communication skills of reading, viewing and listening.

12. Patron engages in a variety of communications situations and attains successive but individual levels of competence. At each level or departure point, the librarian employing interviewing or discussion skills helps the patron access his own level of development.

13. If the patron requires more support in this assessment at any level than can be given by the librarian, referral is made and contact established with a helping consultant, an accepting group, or responsive community movement.

14. Periodically, the patron is encouraged to join with other persons in teams, seminars, or groups for the purpose of developing new ideas or creating new programs. The most important criterion for joining such groups is the prior attainment of suitable skills or knowledge relevant to the activity to be undertaken and the processes of the group. Counseling for group processes is as essential as for content.

15. As the patron develops into a resource person in his own right, the librarian offers him the opportunity of conducting sessions in the library's own programs of communication. Supportive counseling sessions as well as rehearsal demonstrations before and after these leadership situations may be needed.

16. Periodically, the patron should have the opportunity in taking responsibility for the development and conducting of entire programs whether on the library premises, over closed circuit television or through the channels of mass media owned by the library.

17. Eventually, when the patron has developed his ability as one of the communications "elite" he may be able to take leadership in community developement and movements.

Communicator Related Dependent Variables:

1. Dyad Training Needs:
    - Counseling and interviewing
    - Materials composition and therapy
    - Reading, viewing and listening skills advisement
    - Resource tutorials
    - Search and retrieval strategies

2. Group Training Needs:
    - Group dynamics and discussion
    - Educational and program planning
    - Instruction in resource center use
    - Group and meeting organization

3. Community Training Needs:
    - Articulation of concerns and interests
    - Communications leadership
    - Street and floating liaison
    - Community development and organization
    - Media design and coordinating

Agency Related Dependent Variables:

1. Infrastructure axioms
    - Acquire one copy of every human product
    - Organize for access
    - Index for data search and retrieval
    - Promote surprise value of data

2. Communicative standards
    - Continuous audience or market analysis of concerns and interests

Identify and secure or access community and out-of-community resources

Calendar agency and organizational programs, and help improve programs of agencies and organizations.

Make it difficult for people to avoid thinking about personal, group and community issues.

Evaluate the communicative effectiveness of information, library and media centers.

Community Related Dependent Variables:

1. Socioeconomic and cultural factors, e.g. Hall's PMS matrix
2. Coordinating community structures
3. Data and information space

In considering these variables in research design, the point has been made in previous chapters as to the importance of identifying immediate and ultimate objectives. The process is twofold: identifying behavioral outcomes in operational terms, as well as the way outcomes are to be measured. In review, the steps to be taken in the identification and selection of the dependent variable include the following:

Identify the patron behavior to be changed (e.g. search and retrieval performance).

Analyze the patron behavior (e.g. search strategy may be divided into data retrieval, trend identification, logical display, and discovery thinking).

Select the significant specific learner behavior, e.g. the researcher may be particularly interested in data retrieval.

> Define dependent variable operationally and specify how specific behaviors will be measured (e.g. categorize data retrieval inquiries and relate to appropriate data compilations).

When the researcher has selected the dependent variable as well as the specific manner it is to be measured, it becomes necessary to describe the dependent variable in an operational and unambiguous manner so that it can be recognized by another researcher. The following statement may serve as an adequate description of a dependent variable in measurable terms:

> The dependent variable in this study is data retrieval and is defined as the extent to which a subject can categorize data retrieval inquiries within specified time limits. Data retrieval will be measured by a group test administered to all subjects. The instructions in the test are to check one of nine categories which are standard for all inquiries. Thirty seconds will be allowed for each inquiry. The score on this measure will be the number of correct categories identified.

Obviously many research studies are not experimental and many of the variables studied cannot be manipulated. Non-manipulable variables (variates) can be investigated in status (survey) studies, while the manipulable variables which the researcher chooses not to manipulate can be the emphasis in associational (correlational) studies. Both survey and correlational analysis may serve in the process of discovery research. However, correlational analysis has a much greater prognostic value for experimental research. Experimental studies can be used to predict the effect of induced change. Carefully planned manipulation of variables under researcher control constitute the research strategy with the greatest potential for identifying the procedures and programs useful for change agents.

In order to conduct an experimental study the independent variable must not only be identified but its relation to dependent variables must also be specified. There are many variables which can be independent, some of which can be manipulated, others not.

In general, four types of independent variables exist for communications research in information, library and media science: communicative, patron, communicator, environmental (see figure, Independent Variables). Eventually the non-manuipulable independent variables may be analyzed in associational studies whether these be historical or current.

## SOURCES OF INDEPENDENT VARIABLES

| Communicative | Patron | Communicator | Environmental |
|---|---|---|---|
| Materials<br>Inner, outer, package form. | Biological factors<br>maturation<br>health<br>sex | Intelligence and cognitive flexibility | Infrastructure type and network affiliation |
| Agency organization<br>Patron-professional ratio<br>Individual, group, community services | Aptitude and ability: general and specific subject | Professional development and competence | Socioeconomic factors and cultural-peer group values. |
| Communication mode<br>Discovery<br>Expository<br>Programed | Educational and social factors<br>Academic success<br>Family efforts<br>Organizational and group affiliation | Personal development and tolerance of ambiguity | Eg.: Hall's primary message system and matrix |
| Knowledge organization<br>Hierarchial shelf and index arrangement<br>Problem referenced arrangement | | Social development and organizational ability | |

While there may be four general sources of independent variables, some strategy will have to be taken for the identification and selection of an appropriate independent variable. The option of selecting both dependent and independent variables may be the researcher alone or the decision of a research committee. In any event, the researcher has the responsibility to ensure that all significant variables are considered, and to work out procedures for research design and implementation. Review questions may include the following:

> What variables are known to be related to the behavioral outcomes in which I am interested?
>
> How have others defined and described these variables? Has their methodology been sound and comprehensive?
>
> How have the variables been manipulated in previous studies? Have other probable ways been suggested of manipulating them?

Even though sources have been identified for variables in the communicative process, some strategy will have to be developed by the individual researcher. These strategies will include the survey of previous research as well as the analysis of data compilations and banks as well as the results of pilot exploratory studies.

Another possibility includes the observation and analysis of currently operational communications settings in order to determine the events and stimuli which appear to have some effect on the dependent variable. The figure: Identifying Independent Variables may serve to minimize the possibility of missing significant variables or of selecting inappropriate manipulations.

## IDENTIFYING INDEPENDENT VARIABLES

| Potentially Relevant Variable | Variable Status | Potential Relevance |
|---|---|---|
| 1. Communicative Variables | | |
| Materials: inner, outer and package form | All information, library and media centers use form divisions | Understanding of, and ability to use form divisions critical in search and retrieval performance. |
| Agency organization: | | |
| Patron-professional ratio | Varies by type and size of center and document pool | Low ratio provides greater range of successful performance |
| Individual, group and community services | Most centers provide individual services | Higher performance if all types of services available |
| Communication mode: | | |
| Programed | Museums and media centers | Critical for low achievers |
| Discovery | Most centers | Critical for information seekers |
| Expository | Few centers provide | Critical for program and mass communication development |
| Knowledge Organization: | | |
| Hierarchial shelf and index arrangement | Almost "standard" in all centers | Requires at least high school background |
| Problem referenced organization | Few centers provide | Critical for average and low achievers |

| Potentially Relevant Variable | Variable Status | Potential Relevance |
|---|---|---|
| 2. Patron Variables: | | |
| Maturation | Self motivated continuing learners reach higher levels of performance | Performance in search and retrieval varies with maturation |
| Aptitude and ability | Patrons with aptitude and ability reach higher levels of performance | Performance in search and retrieval varies with aptitude |
| Educational and social factors | Patrons with higher educational level reach higher levels of performance | Performance in search and retrieval varies with education |
| 3. Communicator Variables: | | |
| Cognitive flexibility | Patrons helped by skilled problem solvers reach high levels of performance | Performance in search and retrieval varies with problems solving ability of communicator. |
| Tolerance of ambiguity | Patrons prefer assistance from tolerant professionals. | Achievement and motivation vary with degree of pluralism in professional's values. |

| Potentially Relevant Variable | Variable Status | Potential Relevance |
|---|---|---|
| 4. Environmental Variables: | | |
| Infrastructure type and network affiliation | Individual centers not backed up by networks | Use of materials varies inversely with access time. |
| Socioeconomic and cultural factors | Community study and market research not frequently done. | Program impact on target audience determined by context within which communication occurs. |
| 5. Other Variables:<br>... | ... | ... |

Questions For Study And Discussion:

1. Discuss the value of research in information, library and media science as well as indicate those areas of the field most in need of research.

"Kaleidoscope View of Library Research." Wilson Library Bulletin. 41:896-949, May 1967.

Robert H. Muller, "Research Mind in Library Education." Library Journal. 92:1126-29, March 15, 1967.

Douglas M. Knight ed. Libraries at Large. New York, Bowker, 1969.

Sidney G. Tickton ed. To Improve Learning. 2 Vol., New York, Bowker, 1970.

2. Discuss the basic characteristics of research as well as the essential steps in the scientific method. What are the major factors to be considered in the selection of a research proposal.

Bernard S. Phillips. Social Research: Strategy and Tactics. New York, Macmillan, 1966.

Herbert S. Zim, "Science: Process of Discovering New Facts." Drexel Library Quarterly. 5:135-43, July 1969.

Herbert Goldhor. Introduction to Scientific Research in Librarianship. Champaign, Illinois: Illini Union Bookstore, 1969.

3. Discuss the purpose and function of measurement as well as the four types of measurement scales.

Bernard S. Phillips. Social Research: Strategy and Tactics. New York, Macmillan, 1966.

Herbert Goldhor. Research Methods in Librarianship: Measurement and Evaluation. Champaign, Illinois, Illini Union Bookstore, 1967.

4. Consider the nature, function and use of descriptive statistics as well as the commonly associated distributions and measures.

Celeste McCollough. *Statistical Concepts: Program for Self Instruction*. New York, McGraw-Hill, 1963. (especially chapters 6, 8-11).

5. Consider the nature, function and use of inferential statistics as well as the three basic tests of significance.

Celeste McCollough. *Statistical Concepts: Program for Self Instruction*. New York, McGraw-Hill, 1963. (especially chapters 7, 13, 15-19).

6. Discuss the main items to be included in a research proposal and point out the various weaknesses to be avoided.

Gerald R. Smith, "Inadequacies in Research Proposals." *Library Trends*. 13:68-83, July 1964.

J. Krug. "Writing Research Proposals." *ALA Bulletin*. 61:1314-18, December 1967.

7. Identify and discuss the following terms in relation to research design:

judgment
premise
conclusion
reliability
null hypothesis
theory
denotation
connotation
hypothesis
verification
sample
evaluation
induction
deduction
facts
definition
population

David Berlo. *Process of Communication*. New York, Holt, Rinehart and Winston, 1960.

Celeste Mc Collough. *Statistical Concepts: Program for Self Instruction*. New York, McGraw-Hill, 1963. (especially chapters 1-5).

8. Identify and discuss the various methods of research which are commonly employed for the development of hypotheses, the collection and analysis of data, and for the drawing of conclusions.

Bernard S. Phillips. Social Research: Strategy and Tactics. New York, Macmillan, 1966.

Guy Garrison ed., "Research Methods in Librarianship." Library Trends. 13:1-149, July 1964.

Herbert Goldhor. Introduction to Scientific Research in Librarianship. Champaign, Illinois: Illini Union Bookstore, 1969.

Eugene Webb. Unobtrusive Measures: Nonreactive Measures in the Social Sciences. Chicago, Rand McNally, 1966.

Maurice B. Line. Library Surveys. Hamden, Connecticut: Archon Books, 1967.

Pauline V. Young. Scientific Social Surveys and Research 4th ed. Englewood Cliffs, New Jersey: Prentice Hall, 1968.

9. Identify and discuss content analysis and market analysis as research methods as well as the importance of these methods for information, library and media science.

Lester Asheim, "Research in Mass Communications and Adult Reading." Library Trends. 6:120-40, October, 1957.

Richard W. Budd. Content Analysis for Communication. New York, Macmillan, 1967.

Patrick R. Penland, "Content Analysis." Encyclopedia of Library and Information Science. Dekker, 1968-

"Market Analysis" International Encyclopedia of the Social Sciences. New York, Macmillan, 1968.

10. Compare and contrast simulation as a research method with that of systems and model development and analysis.

Bernard S. Phillips. Social Research: Strategy and Tactics. New York, Macmillan, 1966.

Sarane S. Boocock. Simulation Games in Learning. New York, Sage Publications, 1968.

Patrick R. Penland and James G. Williams, "Cybernetic Analysis of Communication Systems." International Conference on Information Science Proceedings, edited by Lydia Vilentchuk. National Center of Scientific and Technological Information. 2 vol. Tel Aviv, Israel, 1972.

11. Identify and characterize the various sources for finding completed research or research in progress in information, library and media science. Either select one research study for analysis and criticism, and/or develop a research proposal of your own in some area of needed research in information, library and media science.

## RESEARCH REPORT ANALYSIS:

Each student will want to report on and analyze, a plan for a research problem in library science. The following outline may be followed in the presentation:

Title (Indicate the title in precise terms).

Selection of the problem.
- Source (Indicate what suggested the topic to you).
- Justification: Justify the selection of the problem by listing reasons with respect to its personal suitability and its general value.
- Scope: Indicate the proposed scope of the study by stating briefly the kind of persons, situations, materials to which the study is to apply.

Procedure of Solution
- Logical analysis: Give in systematic and in logically related arrangement the issues, elements, or subtopics into which the problem may be divided.
- Research procedure: Describe briefly in related fashion the research procedures to be used in attacking the several elements of the problems.
- Data needed: List the specific data, facts, or information to be obtained.
- Procurement of data: Show how each set of data is to be procured, including reference to the means by which personal contacts are to be made.
- Treatment of data: Show how the data are to be treated to be made meaningful as applied to the solution of the problem.
- Assumptions made: List the assumptions or things taken for granted with respect to the several phases of the procedure employed in studying the problem.

Conclusions
- Hypothetical conclusions: Indicate two or three of the conclusions or kinds of results that might grow out of the study.
- Implications of conclusions: Indicate some of the implications of the hypothetical results listed; that is give several inferences bearing on the problem or its probable applications.

Previous studies

Briefly summarize the previous related work done and list bibliographic references to research in the field or to related studies which might serve as models for the solution of the problem at hand.

## Research Proposal Evaluation

1. Methods of Research - Command of proper methods of research or experimentation.

Superior Acceptable Inferior

Comments:

2. Use of Old Material - Command of the literature of the subject.

Superior Acceptable Inferior

Comments:

3. Use of New Material - Industry, skill, and judgement in gathering new material.

Superior Acceptable Inferior

Comments:

4. Originality and insight in interpreting old and new material.

Superior Acceptable Inferior

Comments:

5. Style - clearness, logical arrangement and English style

Superior Acceptable Inferior

Comments:

6. Scope - Adequacy of scope and substance.

| | | |
|---|---|---|
| Superior | Acceptable | Inferior |

Comments:

7. Recommendation - Do you recommend the acceptance of this proposal as a research study.

Remarks:

signature

date

# GLOSSARY

Action Research: Study of a social phenomenon in its usual setting and under typical conditions rather than under controlled experimental conditions. It might be a study of the operational procedures of a home-loan department of an information, library and media center.

Achievement test. A test that measures the extent to which a person has "achieved" something--acquired certain information or mastered certain skills, usually as a result of specific instruction.

Adequate sample. A sample which contains a great enough number of cases (over 30) to have a sampling error within certain desirable limits.

Adverse question. A question that is asked during an interview which tacitly lets the respondent know that nothing will shock or surprise the interviewer, so that he may feel free to express himself without reservation.

Affective behavior. Behavior characterized by an emotional response usually differentiated from cognitive behavior.

Aided-recall technique. The method of getting an interview respondent's views on some issue by asking him to select one of the several prepared opinion statements with which he most nearly agrees.

Alternate-form reliability. The closeness of correspondence or correlation, between results on alternate (i.e. equivalent or parallel) forms of a test: thus, a measure of the extent to which the two forms are consistent or reliable in measuring whatever they do measure, assuming that the examinees themselves do not change in the abilities measured between the two testings.

Analogous practice. An activity similar but not identical to the terminal behavior.

Appropriate practice. An activity identical to that specified or implied by the behavioral objective.

Aptitude. A combination of abilities and other characteristics, whether native or acquired, known or believed to be indicative of an individual's ability to learn in some particular area. Some exclude motivational factors, including interests, from the concept of "aptitude," but the more comprehensive use seems preferable. The layman may think of "aptitude" as referring only to some inborn capacity: the term is no longer so restricted in its psychological or measurement usage.

Arithmetic mean. The sum of a set of scores divided by the number of scores. (Commonly called average, mean.)

Associational study. A study in which the variate is potentially manipulable, but is not manipulated.

Assuming question. An interview question by which the interviewer tacitly allows the respondent to assume that the interviewer is informed about certain facts, even though he may not be.

Average. A general term applied to measures of central tendency. The three most widely used averages are the arithmetic mean, the median, and the mode.

Battery. A group of several tests standardized on the same population, so that results on the several tests are comparable. Sometimes loosely applied to any group of tests administered together, even though not standardized on the same subjects.

Behavior. Individual performance which is directly or indirectly observable.

Behavior mode. The physical characteristics of the learner's response, e.g., written, oral, pointing.

Behavioral objective. The planned result or specified outcome of instruction as it relates to behavior or product.

Bias. A predisposition to evaluate and interpret social phenomena in such a way as to deliberately uphold a preferred frame of reference, and deliberately suppress an opposing one. Some writers speak of "unconscious biases", while others imply that viewing social phenomena from any frame of reference whatever constitutes a bias.

Biased error. An error due to a bias in the selection of data, or in observing them.

Biased sample. A sample which contains a systematic error, and therefore does not represent a given universe of cases.

Cause. The necessary event, or sequence of events which invariably precede a given kind of event: an event or sequence of events which produced a given particular event: an event or factor which contributes, along with other factors, to a given kind of event.

Case Study Method. A method of research in which the significant experiences of a person, or a group of people are recorded and analyzed in order to understand not only the particular case that was studied, but also to gain insights (not generalizations) into a type of social conduct that is common to a whole group or category of people.

Central tendency. The tendency of data to cluster around a mean.

Chi-square test. A test for the goodness or fit of actual data to a hypothetical normal, or chance distribution of data around a line of regression. It is also used as a test of the independence of a collection of data from the principles that were used to classify them.

Closed form questionnaire. A questionnaire from which the respondent chooses one of several prepared alternatives.

Coefficient of correlation (r). A measure of the degree of relationship, or "going-togetherness," between two sets of measures for the same group of individuals. The correlation coefficient most frequently used in test development and educational research is that known as the Pearson (Pearsonian) r, so named for Karl Pearson, originator of the method, or as the product-moment r, to denote the mathematical basis of its calculation. Unless otherwise specified, "correlation" usually means the product-moment correlation coefficient, which ranges from .00, denoting complete absence of relationship, to 1.00, denoting perfect correspondence, and may be either positive or negative.

Coefficient of determination. A figure showing the extent to which a variation in the dependent variable may be explained by changes in the independent variable.

Cognitive behavior. Behavior characterized by a verbal response. Usually differentiated from affective and psychomotor behavior.

Completion item. A test question calling for the completion (filling in) of a phrase, sentence, etc., from which one or more parts have been omitted.

Contingency. The relationship between a specified consequent event and a specified learner behavior.

Contingency management. The arrangement of circumstances so that specified consequent events may be made dependent upon the completion of specified learner behaviors.

Control group. A comparison sample in a controlled experiment which preserves the original conditions to which any changes in the experimental group may be compared.

Correlation. Relationship or "going-togetherness" between two scores or measures: tendency of one score to vary concomitantly with the other, as the tendency of students of high IQ to be above average in reading ability. The existence of a strong relationship -- i.e., a high correlation -- between two variables does not necessarily indicate that one has any causal influence on the other.

Correlational study. A study which yields statements of relationships or association, not statements of causality. This category includes both status and associational studies.

Criterion. A standard by which a test may be judged or evaluated: a set of scores, ratings, etc., that a test is designed to predict or to correlate with.

Criterion measure. An instrument or test designed to measure specified outcomes referenced to predetermined objectives and instruction. It is often used as a "post-test."

Cue. A verbal statement providing the minimum information required by the learner to perform the desired behavior. A component of an instructional specification.

Decile. Any one of the nine percentile points (scores) in a distribution that divide the distribution into ten equal parts: every tenth percentile. The first decile is the 10th percentile, the ninth decile the 90th percentile, etc.

Dependent variable. The variable examined in an experiment to determine how it is affected by manipulation of the independent variable.

Depth interview. An interview which seeks to get not only the replies to certain questions but also some insight about the values and feelings behind them.

Describe. One of the general verb classes used in stating behavioral objectives in which the learner generates and reports the necessary categories of object properties, events, event properties and/or relationships relevant to a designated referent.

Deviation. The amount by which a score differs from some reference value, such as the mean, the norm, or the score on some other test.

Directed interview. An interview method which seeks certain specific data, and which keeps the respondent close to the subject and to the limited number of questions which need to be answered.

Discriminating power. The ability of a test item to differentiate between persons possessing much of some trait and those possessing little.

Distractor. Any of the incorrect choices in a multiple-choice or matching item.

Distribution (frequency distribution). A tabulation of scores from high to low, or low to high, showing the number of individuals that obtain each score or fall in each score interval.

D. K. response. A "don't know" reply to an interview question.

Elicitor. Verbal statement used to elicit or bring out the desired response from the subject.

En-route behavior. A behavior component not normally expressed in the objective statement but found to facilitate the attainment of the terminal behavior.

Equivalent form. Any of two or more forms of a test that are closely parallel with respect to the nature of the content and the difficulty of the items included, and that will yield very similar average scores and measures of variability for a given group.

Experimental study. Any research study in which a manipulable variable is actually manipulated and the defined experimental units are assigned randomly to the treatments.

Extrapolation. In general, any process of estimating values of a function beyond the range of available data.

Factor. In mental measurement, a hypothetical trait, ability, or component of ability, that underlies and influences performance on two or more tests, and hence causes scores on the tests to be correlated. The term "factor" strictly refers to a theoretical variable, derived by a process of factor analysis, from a table of intercorrelations among tests: but it is also commonly used to denote the psychological interpretation given to the variable--i.e., the mental trait assumed to be represented by the variable as verbal ability, numerical ability, etc.

Factor analysis. Any of several methods of analyzing the intercorrelations among a set of variables such as test scores. Factor analysis attempts to account for the interrelationships in terms of some underlying "factors," preferably fewer in number than the original variables: and it reveals how much of the variation in each of the original measures arises from, or is associated with, each of the hypothetical factors. Factor analysis has contributed to our understanding of the organization or components of intelligence, aptitudes, and personality: and it has pointed the way to the development of "purer" tests of the several components.

Forced-choice item. Broadly, any multiple-choice item in which the examinee is required to select one or more of the given choices. The term is best used to denote a special type of multiple-choice item, in which the options, or choices, are (1) of equal "preference value" --i.e., chosen equally often by a typical group, but (2) of differential discriminating ability --i.e., such that one of the options discriminates between persons high and low on the factor that this option measures, while the other options do not.

Genetic study. A research design in which the subjects are observed for a relatively long period of time, perhaps over one generation or more.

Hypothesis. An assertive statement of a tentative relationship between two or more variables.

Identify. One of the verb classes used in stating behavioral objectives in which the learner indicates membership or non-membership of specified objectives or events when the name of the class is given.

Independent variable. The variable that is manipulated in an experimental study.

Interpolation. In general, any process of estimating intermediate values between two known points.

Interval scale. The assignment of numerals to subjects or objects based on equated units of measurement which permit both statements of order and the amount of difference between categories.

Inventory test. As applied to achievement tests, a test that attempts to cover rather thoroughly some relatively small unit of specific instruction or training. The purpose of an inventory test, as the name suggests, is more in the nature of a "stock-taking" of an individual's knowledge or skill than an effort to measure in the usual sense. The term sometimes denotes a type of test used to measure achievement status prior to instruction. Many personality and interest questionnaires are designated "inventories," since they appraise an individual's status in several personal characteristics, or his level of interest in a variety of types of activities.

Item. A single question or exercise in a test.

Item analysis. The process of evaluating single test items by any of several methods. It usually involves determining the difficulty value and the discriminating power of the item, and often its correlation with some criterion.

Level of significance. The statistical probability that chance factors alone produced the observed results.

Limits. Descriptions of the characteristics of correct and plausible but incorrect responses or response choices for the desired behavior as essential information for preparing instructional specifications.

Longitudinal study. A study in which the persons or groups who are studied are observed for a relatively long period of time, e.g. genetic study.

Manipulable variable. A variable having different properties or values to which subjects can be randomly assigned.

Matching item. A test item calling for the correct association of each entry in one list with an entry in a second list.

Mean or arithmetical mean. The average obtained by adding up all the scores or measures of a group and dividing the sum by the number of separate quantities.

Median. The score below which 50 percent of the cases in a score distribution fall. If the shape of the distribution is distorted by the presence of a few aberrant cases of little importance, the median may be a better summary description of the group than the mean. If the distribution is symmetric, the median and mean will be almost identical.

Mode. The score or value that occurs most frequently in a distribution.

Multiple-choice item. A test item in which the examinee's task is to choose the correct or best answer from several given answers, or options.

Multiple-response item. A special type of multiple-choice item in which two or more of the given choices may be correct.

N. The symbol commonly used to represent the number of cases in a distribution, study, etc.

Name. One of general verb classes used in stating behavioral objectives in which the learner supplies the correct verbal label for a referent or set of referents when the label for the referent is not given.

Negative consequence. The presentation of an undesirable condition which increases the probability that the preceding behavior will not occur again under similar conditions.

Nominal scale. The assignment of numerals, letters, or symbols as an identifying label to each subject or object with no ordering of the various categories intended.

Non-manipulable variate. A variable having properties or values to which subjects cannot be randomly assigned.

Non-directed interview. The technique of interviewing in which the informant is allowed to speak on any aspect of a given subject of inquiry that he chooses for as long as he pleases. The aim is usually to determine what the informant thinks is important about the subject of inquiry.

Normal distribution. A distribution of scores or measures that in graphic form has a distinctive bell-shaped appearance, known as a normal curve or normal probability curve. In a normal distribution, scores or measures are distributed symmetrically about the mean, with as many cases at various distances above the mean as at equal distances below it, and with cases concentrated near the average and decreasing in frequency the further one departs from the average, according to a precise mathematical equation.

Norms. Statistics that describe the test performance of specified groups, such as pupils of various ages or grades in the standardization group for a test. Norms are often assumed to be representative of some larger population, as of pupils in the country as a whole. Norms are descriptive of average, typical, or mediocre performance; they are not to be regarded as standards, or as desirable levels of attainment. Grade, age, and percentile are the most common types of norms.

Null hypothesis. A hypothesis that is stated negatively. It is a technique to insure greater objectivity in research. While an investigator is on the lookout for data which will uphold his hypothesis, he is also deliberately sensitive to those data which would contradict his hypothesis and therefore, uphold his null hypothesis.

Objective test. A test in the scoring of which there is no possibility of difference of opinion among scorers as to whether responses are to be scored right or wrong. It is contrasted with a "subjective" test--e.g., the usual essay examination to which different scorers may assign different scores, ratings, or grades.

Open-ended question. A question used in a survey or poll which allows any spontaneous answer from the respondent.

Open-form questionnaire. A questionnaire in which the respondent is asked to write out his responses to the questions asked, instead of selecting from among several previously prepared alternatives to the questions.

Operational definition. A description of a specific observation giving the procedures and conditions under which the observation occurred, as well as the degree of reliability or the chances that another observer under precisely the same conditions and using the same procedures will observe precisely the same phenomena.

Operations analysis. The stage in the product development cycle in which the adequacy of the procedures which were employed in preparing the product are appraised.

Order. One of general verb classes used in stating behavioral objective in which the learner is given specified instructions for arranging two or more class members which are given or must be recalled.

Ordinal scale. The assignment of numerals to subjects or objects which are used to rank order the various categories with respect to some characteristic. The amount of differences between and among categories is disregarded.

Percentile (P). A point (score) in a distribution below which falls the per cent of cases indicated by the given percentile. Thus the 15th percentile denotes the score or point below which 15 per cent of the scores fall. "Percentile" has nothing to do with the per cent of correct answers an examinee has on a test.

Percentile rank. The percent of scores in a distribution equal to or lower than the score corresponding to the given rank.

Performance test. As contrasted with paper-and-pencil test, a test requiring motor or manual response on the examinee's part, generally but not always involving manipulation of concrete equipment or materials. "Performance test" is also used in another sense, to denote a test that is actually a work-sample, and in this sense it may include paper-and-pencil tests, as, for example, a test in accountancy, or in taking shorthand, or in proofreading, where no materials other than paper and pencil may be required, but where the test response is identical with the behavior about which information is desired.

Personality test. A test intended to measure one or more of the non-intellective aspects of an individual's mental or psychological make-up, called personality inventories or adjustment inventories which seek to measure a person's status on such traits as dominance, sociability, introversion, etc., by means of self-descriptive responses to a series of questions: rating scales which call for rating, by one's self or another, of the extent to which a subject possesses certain characteristics: situation tests in which the individual's behavior in simulated life-like situations is observed by one or more judges, and evaluated with reference to various personality traits: and opinion or attitude inventories. Some writers also classify interest inventories as personality tests.

Population. That portion of a defined universe to which a researcher has access, e.g., all adults who cannot read beginning word elements.

Positive consequence. The presentation of a condition which increases the probability that the preceding behavior will occur again under similar conditions.

Power test. A test intended to measure level of performance rather than speed of response: hence one in which there is either no time limit or a very generous one.

Premack principle. An assertion that of any two activities in which a person is freely engaged, the one in which he is involved more frequently can function as a positive consequence for the other one.

Practice effect. The influence of previous experience with a test on a later administration of the same test or a similar test: usually, an increase in the score on the second testing, attributed to increased familiarity with the directions, kinds of questions, etc. Practice effect is greatest when the interval between testings is small, when the materials in the two tests are very similar, and when the initial test-taking represents a relatively novel experience for the subjects.

Profile. A graphic representation of the results on several tests, for either an individual or a group, when the results have been expressed in some uniform or comparable terms. This method of presentation permits easy identification of areas of strength or weakness.

Projective technique (projective method). A method of personality study in which the subject responds as he chooses to a series of stimuli such as ink-blots, pictures, unfinished sentences, etc. So called because of the assumption that under this free-response condition the subject "projects" into his responses manifestations of personality characteristics and organization that can, by suitable methods, be scored and interpreted to yield a description of his basic personality structure.

Prognosis (prognostic) test. A test used to predict future success or failure in a specific subject or field.

Quartile. One of the three points that divide the cases in a distribution into four equal groups. The lower quartile, or 25th percentile, sets off the lowest fourth of the group: the middle quartile is the same as the 50th percentile, or median: and the third quartile, or 75th percentile, marks off the highest fourth.

Questionnaire. A series of written questions given to a respondent to answer without any assistance or intermediation from the issuing source.

Random sample. A sample of the members of a population drawn in such a way that every member of the population has an equal chance of being included--that is, drawn in a way that precludes the operation of bias or selection. The purpose in using a sample thus free of bias is, of course, that the sample be fairly "representative" of the total population, so that sample findings may be generalized to the population. A great advantage of random samples is that formulas are available for estimating the expected variation of the sample statistics from their true values in the total population: in other words, we know how precise an estimate of the population value is given by a random sample of any given size.

Randomization. The process of selecting a portion of a population in which all members of the population have equal probabilities of being included in the sample.

Range. The difference between the lowest and highest scores obtained on a test by some group.

Raw score. The first quantitative result obtained in scoring a test. Usually the number of right answers, number right minus some fraction of number wrong, time required for performance, number of errors, or similar direct, unconverted, uninterpreted measure.

Reactive measures. A measurement procedure which by its use alone, may produce a change in the learner's behavior.

Readiness test. A test that measures the extent to which an individual has achieved a degree of maturity or acquired certain skills or information needed for undertaking successfully some new learning activity.

Recall item. An item that requires the examinee to supply the correct answer from his own memory or recollection, as contrasted with a recognition item, in which he need only identify the correct answer.

Recognition item. An item requiring the examinee to recognize or select the correct answer from among two or more given answers.

Regular interval sample. A random sample that is derived by using a systematic method of drawing cases from a list, as for example, drawing every tenth case in a universe to form the sample.

Reinforcement. An event which increases the probability that the behavior which just occurred will reoccur under similar conditions.

Reliability. The extent to which a test is consistent in measuring whatever it does measure: dependability, stability, relative freedom from errors of measurement. Reliability is usually estimated by some form of reliability coefficient or by the standard error of measurement.

Reliability coefficient. The coefficient of correlation between two forms of a test, between scores on repeated administrations of the same test, or between halves of a test, properly corrected. These three coefficients measure somewhat different aspects of reliability but all are properly spoken of as reliability coefficients.

Reliability of data. The extent to which the data that were gathered about a social phenomenon accurately reflect the true social phenomenon.

Representative sample. A sample that corresponds to or matches the population of which it is a sample with respect to characteristics important for the purposes under investigation--e.g., in an achievement test norm sample, proportion of pupils from each state, from various regions, from segregated and non-segregated schools, etc.

Research design. The overall plan or strategy for the conduct of a study. In the statistical sense, it details the population specification, sampling plan, and data collection procedures.

Sample. A portion (usually a small fraction) of the population which has the same distribution of essential characteristics as the population.

Sampling. The technique of drawing a few data from a large universe of data in order to study something about the universe.

Sampling error. The degree to which an actual sample which was designed to be a representative sample digresses from an ideal representative sample of that universe.

Sampling plan. A statement which specifies the number of subjects selected, basis for selection, and controls used.

Sample reliability. The degree to which a given sample reflects accurately the universe of cases from which it was drawn.

Sieve question. A question in an interview which reveals to an interviewer whether or not the respondent knows enough to answer certain other questions in the interview.

Skewness. The tendency of a distribution to depart from symmetry or balance around the mean.

Sociometry. Measurement of the interpersonal relationship prevailing among the members of a group. By means of sociometric devices, e.g., the sociogram, an attempt is made to discover the patterns of choice and rejection among the individuals making up the group--which ones are chosen most often as friends or leaders ("stars"), which are rejected by others ("isolates"), how the group subdivides into clusters or cliques, etc.

Spearman-Brown formula. A formula giving the relationship between the reliability of a test and its length. The formula permits estimation of the reliability of a test lengthened or shortened by any amount, from the known reliability of a test of specified length. Its most common application is in the estimation of reliability of an entire test from the correlation between two halves of the test (split-half reliability).

Split-half coefficient. A coefficient of reliability obtained by correlating scores on one half of a test with scores on the other half. Generally, but not necessarily, the two halves consist of the odd-numbered and the even-numbered items.

Standard deviation (S.D.). A measure of the variability or dispersion of a set of scores. The more the scores cluster around the mean, the smaller the standard deviation.

Standard error (S.E.). An estimate of the magnitude of the "error of measurement" in a score--that is, the amount by which an obtained score differs from a hypothetical true score. The standard error is an amount such that in about two-thirds of the cases the obtained score would not differ by more than one standard error from the true score. The probable error (P.E.) of a score is a similar measure, except that in about half the cases the obtained score differs from the true score by not more than one probable error. The probably error is equal to about two-thirds of the standard error. The larger the probable or the standard error of a score, the less reliable the measure.

Standardized test (standard test). A systematic sample of performance obtained under prescribed conditions, scored according to definite rules, and capable of evaluation by reference to normative information. Some writers restrict the term to tests having the above properties, whose items have been experimentally evaluated, and/or for which evidences of validity and reliability are provided.

Stanine. One of the steps in a nine-point scale of normalized standard scores. The stanine (short for standard-nine) scale has values from 1 to 9, with a mean of 5, and a standard deviation of 2.

Statement of association. An indication of the extent of relationship between the variate and the criterion variable.

Statement of causality. An indication of the effect of induced change in the variate on the dependent variable.

Status study. A research study in which the variate is non-manipulable.

Stimulus. A specified environmental event.

Substantive hypothesis. An "if--then" proposition which posits what is likely to occur if various conditions are evoked.

Survey test. A test that measures general achievement in a given subject or area, usually with the connotation that the test is intended to measure group status, rather than to yield precise measures of individuals.

Target population. A group to whom the results of research and development activities are directed.

Terminal behavior. The desired learner behavior or end product for any one unit of behavior or instruction.

Test-retest coefficient. A type of reliability coefficient obtained by administering the same test a second time after a short interval and correlating the two sets of scores.

Treatment. A manipulated variable in an experiment.

True-false item. A test question or exercise in which the examinee's task is to indicate whether a given statement is true or false.

True score. A score entirely free of errors of measurement. True scores are hypothetical values never obtained by testing, which always involves some measurement error. A true score is sometimes defined as the average score of an infinite series of measurements with the same or exactly equivalent tests, assuming no practice effect or change in the examinee during the testings.

Unobrusive measure. Criterion data which does not require the cooperation of the respondent.

Validity. The extent to which a test does the job for which it is used. Validity, thus defined, has different connotations for various kinds of tests and, accordingly, different kinds of validity evidence are appropriate for them.

Variate. Any common characteristic or experience of the subjects that serves as the basis for combining them to form the group or groups being studied, e.g., high school graduation, sex, text assigned, curriculum configuration.

Vital statistics. The raw figures on demographic data such as births, and deaths in a given population.

## GENERAL REFERENCES

T.W. Anderson, *An Introduction to Multivariate Statistics*. New York: John Wiley & Sons, Inc., 1958.

Donald Ary. *Introduction to Research in Education*. N.Y., Holt, Rinehart, and Winston, 1972.

R.L. Baker and R.E. Schutz, (eds.) *Instructional Product Development*. New York: American Book Company and Van Nostrand Reinhold Company, 1971.

Robert L. Baker. *Instructional Product Research*. N.Y., American Book Co., 1972.

Hubert M. Blalock. *Social Statistics*. N.Y., McGraw-Hill, 1960.

John W. Bowers. *Designing the Communication Experiment*. N.Y., Random House, 1970.

Ira R. Buchler. *Game Theory in the Behavioral Sciences*. University of Pittsburgh Pr., 1969.

D.T. Campbell and J.C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," in N.L. Gage (ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, 1963, chap. 5.

F.S. Chapin, *Experimental Designs in Sociological Research*. New York: Harper & Row, Publishers, Incorporated, 1947; revised edition, 1955.

Alphonse Chapins. *Research Techniques in Human Engineering*. Baltimore, John Hopkins Pr., 1959.

William G. Cochran, "The Design of Experiments," in *International Encyclopedia of the Social Sciences*. New York: The Macmillan Company, 1968, Vol. 5, 245-254.

William G. Cochran. *Experimental Designs*. N.Y., Wiley, 1957.

William W. Cooley. *Multivariate Procedures for the Behavioral Sciences*. N.Y., Wildy, 1962.

D.R. Cox. *The Planning of Experiments*. New York: John Wiley & Sons, Inc., 1958.

J.R. Davitz and L.J. Davitz, *Guide for Evaluating Research Plans in Education and Psychology*. New York: Columbia University, Teachers College, 1967.

O.E. Dial. *Computer Programming and Statistics for Basic Research*. New York: Van Nostrand Reinhold Co., 1968.

Sanford M. Dornbusch. *Primer of Social Statistics*. N.Y., McGraw-Hill, 1955.

P.H. Dubois. *Multivariate Correlation Analysis*. New York: Harper & Row, Publishers, Incorporated, 1957.

M.D. Dunnette. "Fads, Fashions, and Folderol in Psychology," *American Psychologist*, 21(1966).

C. Eisenhart, et al. *Selected Techniques of Statistical* Analysis. New York: Dover Publications, Inc., 1970.

Educational Resources Information Center. *How to Use ERIC*. Washington, D.C.: U.S. Office of Education, National Center for Educational Communication, 1970.

Allen L. Edwards, *Experimental Design in Psychological Research*. 3rd edition. New York: Holt, Rinehart and Winston, 1968.

Allen L. Edwards. *Statistical Methods*. 2 ed. N.Y., Holt, Rinehart, and Winston, 1967.

P. Fellin, T. Tripodi, and H.J. Meyer. *Exemplars of Social Research*. Itasca, Ill.: F.E. Peacock Publishers, Inc., 1969.

Leon Festinger. *Research Methods in the Behavioral Sciences*. N.Y., Holt, Rinehart, and Winston, 1953.

J.L. Finan. "The System Concept as a Principle of Methodological Design," in R.M. Gagne (ed.), *Psychological Principles in System Development*. New York: Holt, Rinehart, and Winston, Inc., 1962.

D.J. Fox. *The Research Process in Education*. New York: Holt, Rinehart and Winston, Inc., 1969.

E. Fry. "A Classification of Variables in a Programmed Learning Situation," in J.P. De Cecco (ed.), *Educational Technology*. New York: Holt, Rinehart and Winston, Inc., 1964.

Wilson Gee. *Social Science Research Methods*. N.Y., Appleton-Century-Crofts, 1950.

T.F. Gilbert. "A Structure for a Coordinated Research and Development Laboratory," in J.P. De Cecco (ed.), *Educational Technology*. New York: Holt, Rinehart and Winston, Inc., 1964.

T.K. Glennan, Jr. "Issues in the choice of development policies," in T. Marschak, T.K. Glennan, Jr., and R. Summers, Strategies for R and D. New York: Springer-Verlag, 1967, Chapter 2.

William J. Goode. Methods in Social Research. N.Y., McGraw-Hill, 1952.

Charles Y. Glock. Survey Research in the Social Sciences. N.Y., Russell Sage, 1967.

E.G. Guba. "Guides for the Writing of Proposals," in J.A. Culbertson and S.P. Heneley (eds.) Educational Research: New Perspectives. Danville, Ill.: The Interstate Printers & Publixhers, Inc., 1963.

J.P. Guilford. Fundamental Statistics in Psychology and Education. 4th edition. New York: McGraw-Hill, 1965.

H. Harmon. Modern Factor Analysis. Chicago: University of Chicago Press, 1960.

J.E. Hill and A. Kerber. Models, Methods, and Analytical Procedures in Educational Research. Detroit: Wayne State University Press, 1967.

R. Hyman and B. Anderson. "Solving Problems," International Science and Technology, September, 1965, 36-41.

Palmer O. Johnson. Introduction to Statistical Methods. N.Y., Prentice-Hall, 1953.

A. Kaplan. The Conduct of Inquiry. San Francisco: Chandler Publishing Company, 1964.

F.N. Kerlinger, Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1964.

D.R. Krathwohl. How to Prepare a Research Proposal. Syracuse, N.Y.: Syracuse University Bookstore, 1965.

Daniel Lerner. Evidence and Inference. Glencoe, Illinois, Free Pr., 1958.

S. Levine and F.F. Elzey. A Programmed Introduction to Research. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1968.

E.F. Lindquist. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1953.

C.M. Lindvall. Defining Educational Objectives. Pittsburgh: University of Pittsburgh Press, 1964.

P.R. Lohnes and W.W. Cooley. Introductory Statistical Procedures: With Computer Exercises. New York: John Wiley & Sons, Inc., 1968.

William N. McPhee. Formal Theories of Mass Behavior. Glencoe, Illinois, Free Press, 1963.

R.R. Nelson, J. Peck, and E.D. Kalachek. Technology, Economic Growth, and Public Policy. Washington, D.C.: The Brookings Institution, 1967.

John G. Peatman. Introduction to Applied Statistics. N.Y., Harper and Row, 1963.

Bernard S. Phillips. Social Research: Strategy and Tactics. 2 ed., N.Y.: Macmillan, 1971.

J.W. Popham. Educational Statistics: Use and Interpretation. New York: Harper & Row, Publishers, Incorporated, 1967.

W.S. Ray. An Introduction to Experimental Design. New York: The Macmillan Company, 1960.

Walter A. Rosenblith, ed. Sensory Communication. M.I.T. Press, 1961.

Gilbert Sax. Empirical Foundations of Educational Research. Englewood Cliffs, N.S.: Prentice-Hall, 1968.

R.E. Schutz, E.B. Page, and J.C. Stanley. "Experimental Design in Educational Media Research," in Curriculum Guide for a Course in Educational Media Research. Washington, D.C.: U.S. Office of Education, NDEA Title VII Project B-236, 1962.

C. Selltiz, M. Johoda, M. Deutsch, and S.W. Cook. Research Methods in Social Relations. New York: Henry Holt and Company, Inc., 1951.

M. Sidman. Tactics of Scientific Research. New York: Basic Books, Inc., 1960.

S. Siegel. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

T.D. Sterling and S.V. Pollack. Introduction to Statistical Data Processing. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968.

M.M. Tatsuoka and D.V. Tiedeman. "Statistics as an Aspect of Scientific Method in Research on Teaching," in N.L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally & Company, 1963.

T. Tripodi, P. Fellin, and H.J. Meyer. The Assessment of Social Research. Itasca, Ill.: F.E. Peacock Publishers, Inc., 1969.

D.B. Van Dalen and W.J. Meyer. Understanding Educational Research. New York: McGraw-Hill Book Company, 1966.

D.J. Veldman. FORTRAN Programming for the Behavioral Sciences. New York: Holt, Rinehart and Winston, Inc., 1967.

Helen M. Walker. Statistical Inference. N.Y., Holt, Rinehart, and Winston, 1953.

Eugene J. Webb. Unobtrusive Measures. Chicago, Rand McNally, 1966.

Eric Weiss (ed.) Computer Usage Fundamentals. New York: McGraw-Hill Book Company, 1969.

Frederick F. Whitney. Elements of Research. 3 ed. N.Y.: Prentice-Hall, 1950.

W. Wiersma. Research Methods in Education: An Introduction. Philadelphia: J.B. Lippincott Company, 1969.

D.J. Willower. "Concept Development and Research," in J.A. Culbertson and S.P. Heieley (eds.), Educational Research: New Perspectives. Danvill, Ill.: The Interstate Printers & Publishers, Inc., 1963.

B.J. Winer. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

Pauline V. Young. Scientific Social Surveys and Research. 4 ed. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Hans Zeisel. Say it with Figures. 4 ed. N.Y., Harper & Row, 1957.